NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through National Quality Forum's (NQF) Consensus Development Process (CDP). The information submitted by the measure developers/stewards is included after the *Brief Measure Information* and *Preliminary Analysis* sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3713e

Corresponding Measures:

Measure Title: Hospital Harm - Acute Kidney Injury

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: This electronic clinical quality measure (eCQM) assesses the proportion of inpatient hospitalizations for patients 18 years of age or older who have an acute kidney injury (stage 2 or greater) that occurred during the encounter. Acute kidney injury (AKI) stage 2 or greater is defined as a substantial increase in serum creatinine value, or by the initiation of kidney dialysis (continuous renal replacement therapy (CRRT), hemodialysis or peritoneal dialysis).

1b.01. Developer Rationale:

The goal of this acute kidney injury (AKI) electronic clinical quality measure (eCQM) is to improve patient safety and prevent patients from developing moderate-to-severe AKI (i.e., stage 2 or greater) during their hospitalization. Early identification and management of at-risk patients is critical, as there is no specific treatment to reverse acute kidney injury once occurred (KDIGO, 2012). The Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guidelines for AKI have cited serum creatinine as an acceptable proxy for defining and monitoring AKI, and have provided detailed clinical guidelines to evaluate and monitor patients at-risk of kidney damage (Lopes & Jorge, 2013; KDIGO, 2012).

This eCQM defines the harm of AKI as patients who have a substantial increase in their serum creatinine level, or have to initiate dialysis within 48 hours or more after the start of the hospitalization. We define a substantial increase in serum creatinine as a rise of at least at least 2.0 times higher than the lowest prior serum creatinine value, and the increased value is greater than the highest sex-specific normal value for serum creatinine. This eCQM uses a definition of AKI that is consistent with the definition presented in the KDIGO clinical practice guidelines for Stage 2 or greater (KDIGO, 2012). An increase in serum creatinine indicates a reduction in kidney function, sometimes damaging the kidneys so dialysis is required, and is also associated with an increased risk of mortality (KDIGO, 2012). AKI can cause direct patient harm and symptoms associated with volume overload, electrolyte disorders, uremic complications, and drug toxicity (Lopes & Jorge, 2013; KDIGO, 2012; Hoste & De Corte, 2011). AKI has also been associated with longer term harmful outcomes, such as increased odds of death, increased length of hospital stay, and approximately $7,500 in excess hospital costs (Chertow et al., 2005). This eCQM also captures the need to initiate dialysis after 48 hours of hospital care, as one study found that patients who are treated with renal replacement therapy (dialysis) in the

intensive care unit (ICU) still have an extremely high mortality rate of 50-60% (Hoste & Schurgers, 2008). The desired outcome of this eCQM is to reduce AKI and dialysis initiation rates.

While not all instances of acute kidney injury (AKI) are avoidable and AKI may be due to the natural progression of underlying illness or a complication of a necessary treatment such as chemotherapy, a proportion of AKI cases are preventable and/or treatable at an early stage to improve outcomes. The 2012 Kidney Disease: Improving Global Outcomes (KDIGO) guidelines suggest careful management of hemodynamic status, fluid balance, and vasoactive medications, along with avoidance of nephrotoxic exposures and drug dose adjustment, for the prevention and early treatment of acute kidney injury (KDIGO, 2012). Specific KDIGO recommendations (level 1 = recommended, level 2 = suggested; A=high quality evidence, B=moderate quality evidence, C=low quality evidence) include:

"3.1.1: In the absence of hemorrhagic shock, we suggest using isotonic crystalloids rather than colloids (albumin or starches) as initial management for expansion of intravascular volume in patients at risk for AKI or with AKI. (2B)

3.1.2: We recommend the use of vasopressors in conjunction with fluids in patients with vasomotor shock with, or at risk for, AKI. (1C)

3.1.3: We suggest using protocol-based management of hemodynamic and oxygenation parameters to prevent development or worsening of AKI in high-risk patients in the perioperative setting (2C) or in patients with septic shock (2C).

3.3.1: In critically ill patients, we suggest insulin therapy targeting plasma glucose 110–149 mg/dl (6.1–8.3 mmol/l). (2C)

3.3.2: We suggest achieving a total energy intake of 20–30 kcal/kg/d in patients with any stage of AKI. (2C)

3.8.1: We suggest not using aminoglycosides for the treatment of infections unless no suitable, less nephrotoxic, therapeutic alternatives are available. (2A)

3.8.2: We suggest that, in patients with normal kidney function in steady state, aminoglycosides are administered as a single dose daily rather than multiple-dose daily treatment regimens. (2B)

3.8.3: We recommend monitoring aminoglycoside drug levels when treatment with multiple daily dosing is used for more than 24 hours. (1A)

3.8.4: We suggest monitoring aminoglycoside drug levels when treatment with single-daily dosing is used for more than 48 hours. (2C)

3.8.5: We suggest using topical or local applications of aminoglycosides (e.g., respiratory aerosols, instilled antibiotic beads), rather than intravenous application, when feasible and suitable. (2B)

3.8.6: We suggest using lipid formulations of amphotericin B rather than conventional formulations of amphotericin B. (2A)

3.8.7: In the treatment of systemic mycoses or parasitic infections, we recommend using azole antifungal agents and/or the echinocandins rather than conventional amphotericin B, if equal therapeutic efficacy can be assumed. (1A)

4.3.2: We recommend using either iso-osmolar or low-osmolar iodinated contrast media, rather than high-osmolar iodinated contrast media in patients at increased risk of CI-AKI. (1B)

4.4.1: We recommend intravenous volume expansion with either isotonic sodium chloride or sodium bicarbonate solutions, rather than no intravenous volume expansion, in patients at increased risk for CI-AKI. (1A)

4.4.2: We recommend not using oral fluids alone in patients at increased risk of CI-AKI. (1C)"

This eCQM focuses on stage 2 or higher AKI to encourage entities to identify high-risk patients, and to diagnose AKI at its earliest stage (stage 1), in order to implement interventions to prevent progression. Accurately monitoring the rate at which AKI occurs in the hospital setting will allow hospitals to improve quality and reduce AKI harm rates. Several studies identified through systematic literature searches

developed or evaluated the effectiveness of acute kidney injury electronic alert systems (Selby et al., 2012; Ahmed et al., 2015; Porter et al., 2014; Wilson et al., 2014; Kirkendall et al., 2014; Cho et al., 2012). These studies used data elements for defining AKI that were already present and populated in the EHR. For acute kidney injury diagnosis, all except two were limited to using serum creatinine levels, suggesting that this is the most reliable and consistently available electronic data element for defining acute kidney injury.

---

sp.12. Numerator Statement:

Inpatient hospitalizations for patients who develop acute kidney injury (AKI) (stage 2 or greater) during the encounter, as evidenced by:

- A subsequent increase in serum creatinine value at least 2 times higher than the lowest serum creatinine value, and the increased value is greater than the highest sex-specific normal value for serum creatinine; or
- Kidney dialysis (continuous renal replacement therapy (CRRT), hemodialysis or peritoneal dialysis) initiated 48 hours or more after the start of the encounter.

sp.14. Denominator Statement: Inpatient hospitalizations for patients 18 years of age or older at the start of the encounter without a diagnosis of obstetrics, with a length of stay of 48 hours or longer who had at least one serum creatinine value after 48 hours from the start of the encounter.

sp.16. Denominator Exclusions:

Denominator exclusions are as follows:

- Inpatient hospitalizations for patients with an increase in serum creatinine value of at least 0.3 mg/dL between the index serum creatinine and a subsequent serum creatinine taken within 48 hours of the encounter start (indicating AKI present on admission).
- Inpatient hospitalizations for patients with the index estimated glomerular filtration rate (eGFR) value of <60 mL/min within 48 hours of the encounter start (indicating chronic kidney disease, stage 3a or greater, present on admission).
- Inpatient hospitalizations for patients who have less than two serum creatinine results within 48 hours of the encounter start (indicating that the hospital stay was too short to diagnose AKI).
- Inpatient hospitalizations for patients who have kidney dialysis (CRRT, hemodialysis or peritoneal dialysis) initiated within 48 hours of the encounter start (indicating end stage renal disease, a severe acute metabolic derangement, or AKI present on admission).
- Inpatient hospitalizations for patients with at least one specified diagnosis present on admission that puts them at extremely high risk for AKI:
    - Hemolytic Uremic Syndrome (HUS)
    - Large Body Surface Area (BSA) Burns
    - Traumatic Avulsion of Kidney
    - Rapidly Progressive Nephritic Syndrome
    - Thrombotic Thrombocytopenic Purpura
- Inpatient hospitalizations for patients who have at least one specified procedure during the encounter that puts them at extremely high risk for AKI:
    - Extracorporeal membrane oxygenation (ECMO)
    - Intra-Aortic Balloon Pump
    - Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA)
    - Nephrectomy

Measure Type: Outcome

sp.28. Data Source:

Electronic Health Records

sp.07. Level of Analysis:

Facility

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a ***health outcome*** measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance can be used, assuming the data are from a robust number of providers and the results are not subject to systematic bias. For measures derived from a patient report, the evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

The developer provides the following description for this measure:

- This is a new electronic clinical quality measure (eCQM) outcome measure at the facility level that assesses the proportion of inpatient hospitalizations for patients 18 years of age or older who have an acute kidney injury (stage two or greater) that occurred during the encounter.
- The developer provides a logic model that depicts that high risk individuals should receive a kidney health assessment (which includes AKI history, blood pressure, chronic kidney disease and serum Creatinine level, drug list, and urine dipstick) and depending on the results of the assessment a kidney health response is initiated. The kidney health response includes medication review and adjustment, minimize nephrotoxic exposures, message healthcare team to alert them to the high risk of AKI and to optimize the patient's volume status and hemodynamic parameters, and monitor for AKI and its consequences. The logic model attests that if the kidney health response is implemented, it is expected to lead to primary prevention of AKI, prevention of progression from stage one to stage two, reduction of risk of dialysis initiation, and improved long-term outcomes.

Summary:

- To demonstrate how the target population values the measured outcome, the developer stated that groups of experts convened in 2019 to discuss evidence and determine recommendations for AKI quality indicators and care processes as well as understand how treatment of AKI has advanced since the 2012 KDIGO guidelines.
    - The experts concluded that consensus of clinical opinion and several recent trials support evidence-based fluid management and drug stewardship to reduce AKI and provided consensus recommendation statements on AKI treatment.
- The developer cited a 2009 meta-analysis of randomized controlled trials on the effects of perioperative hemodynamic goal-directed therapy for adult surgical patients that included 4,220 patients and found that the odds of postoperative acute renal injury were significantly reduced by hemodynamic optimizations when compared to the control group.

- The developer further stated that evidence has confirmed the effectiveness of the 2012 KDIGO recommendations in preventing AKI through two small single center randomized controlled trials called Prevention of AKI (PrevAKI) and Biomarker Guided Intervention to Prevent AKI after Major Surgery (BIGpAK)
    - The PrevAKI trial included 276 high-risk adults undergoing cardiac surgery with cardiopulmonary bypass. Patients were either allocated to receive the KDIGO bundle or to the control group.
        - The KDIGO bundle group has significantly lower rates of moderate and severe AKI compared to the control group.
    - The BigpAK trial included 121 patients with increased AKI risk after major abdominal surgery. Patients were either allocated to receive the biomarker-guided KDIGO care bundle or to the standard care group.
        - The biomarker-guided KDIGO care bundle group significantly reduced incidence of moderate and severe AKI compared to the standard care group.

Question for the Standing Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance From the Evidence Algorithm

Measure assesses performance on a health outcome (Box 1) à Relationship between the health outcome and at least one healthcare action (Box 2) à Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer used data from the 20 hospitals that participated in the measure testing using data from the full 2020 calendar year.
    - The developer stated that the observed performance rate in acute kidney injury ranged from a minimum of 0.76 percent to a maximum of 4.43 percent per 100 qualified admissions.
    - The weighted average measure rate was 1.52 percent per 100 qualified inpatient admissions. The interquartile range was 0.66 percent unadjusted and 0.84 percent adjusted.
- The developer also stated that at least one prior study of critically ill patients admitted to the intensive care units at six hospitals in four countries applied the KDIGO criteria to estimate variation in the incidence of stage one or greater AKI.
    - Of the 15,132 patients, 32 percent developed AKI based on serum creatinine criteria, but this risk varied widely from 14.6 to 43.8 percent. The developer further stated that after adjusting for differences in age, sex, and severity of illness the odds ratio for AKI continue to vary across centers (2.57 – 6.04).
    - The developer also notes, based on several published studies, that the incidence of AKI in hospitalized patients is comparable to the rates of severe sepsis and acute lung injury, and that AKI may result in the need for dialysis as well as increased risk of mortality.

Disparities

- The developer presented the rate of AKI per 100 denominator encounters for different subgroups:
    - For age, the 65 plus group had the highest rate of AKI in the Meditech system (4.21 percent) while ages 36-64 had the highest rate of AKI in the Cerner system (1.38 percent).
    - For sex, in both the Meditech and Cerner systems males had a higher rate of AKI at 4.04 percent and 1.42 percent respectively, compared to females who had a rate of 3.74 percent in Meditech and 1.22 percent in Cerner.
    - For race, in the Meditech system Other had the highest rate of AKI (6.25 percent) compared to White (3.20 percent) and Black or African American (5.25 percent). In the Cerner system, Other had the highest rate of AKI (1.57 percent) compared to White (1.27 percent), Black or African American (1.22 percent), and Unknown (1.40 percent).
    - For ethnicity, in both the Meditech and Cerner systems Hispanic had the highest rate of AKI at 5.21 percent and 1.46 percent respectively.
    - For payer type, in the Meditech system, Other had the highest rate of AKI (5.00 percent) compared to Medicare (4.10 percent), Medicaid (3.83 percent), private insurance (3.73 percent), and unknown (3.81 percent). In the Cerner system, private had the highest rate of AKI (1.43 percent) compared to Medicare (1.38 percent), Medicaid (1.33 percent), self-pay or uninsured (1.13 percent), and other (0.88 percent).

Questions for the Standing Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by the Scientific Methods Panel (SMP)? ☐ **Yes** ☒ **No**

Evaluators: Staff

2a. Reliability: Specifications and Testing

2a1. Specifications require the measure, as specified, to produce consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented.

- The submitted measure specification follows established technical specifications for electronic clinical quality measures (eCQMs) (Quality Data Model [QDM], health quality measure format [HQMF], and Clinical Quality Language [CQL]) as indicated in subcriterion 2a1.
- The submitted measure specification is fully represented and is not hindered by any limitations in the established technical specifications for eCQMs.

2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.

Specifications:

- Measure specifications are clear and precise.
- eCQMs as specified using the latest industry-accepted eCQM technical specifications: HQMF, QDM, CQL, and value sets vetted through the National Library of Medicine's (NLM) Value Set Authority Center (VSAC).

Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:
    - Testing was conducted using Adams' signal-to-noise ratio (SNR) and an intra-class correlation coefficient (ICC) using a split-half sample approach on electronic health records from 58,936 denominator encounters across 20 hospitals with bed sizes ranging from 25 to over 499 from January 1, 2020, to December 31, 2020.
        - The developer reported that the SNRs ranged from 0.20 to 0.97 with a mean of 0.84 and median of 0.91.
        - The developer reported the 100 estimated ICCs for the observed measure rates had a median ICC of at 1.0 and no simulations generated median ICCs below 0.99. The mean ICC values ranged from 0.25 to 0.91.
        - The developer reported the 100 estimated ICCs for the adjusted measure rates had a median ICC of 0.99 and all but two simulations generated median ICC values over 0.95. The median value of the mean ICC for risk-adjusted measure rates was 0.62.

Questions for the Standing Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*

Guidance From the Reliability Algorithm

Specifications precise, unambiguous, and complete (Box 1) → Empirical reliability testing conducted statistical tests as the measure is specified (Box 2) → Empirical reliability testing conducted at the accountable entity level (box 4) → Method described for reliability testing was appropriate (Box 5) → High certainty or confidence that accountable entity levels are reliable (Box 6a) → High

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity: Validity Testing; Exclusions; Risk Adjustment; Meaningful Differences; Comparability; Missing Data

2b2. Validity testing should demonstrate that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Validity Testing

- Validity testing conducted at the Patient/Encounter Level:
    - The developers compared data exported from the electronic health record (EHR) to data manually abstracted from medical charts for a subsample of the measure's initial population.
    - The developer then calculated the rate of missingness for all critical data elements (results are reviewed in the missing data section below), the percentage match between data sources, positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity. They then assessed whether excluded cases for EHR data truly met the intent for exclusion.
        - Match/Agreement
            - For each of the 17 Cerner sites, the data elements had a 100 percent match rate.
            - Site 18 and 19, the first and second Meditech site, had a 100 percent match rate for all data elements.
            - Site 20, the third Meditech site, had six data elements that had a 100 percent match rate, and one data element had a 95 percent match rate.

        - PPV/NPV/Sensitivity/Specificity
            - For the 17 Cerner sites, PPV ranged from 90.6 percent to 100 percent, NPV ranged from 94 percent to 100 percent, sensitivity ranged from 79 percent to 100 percent, and specificity ranged from 93.9 percent to 100 percent.
            - For site 18, the first Meditech site, PPV ranged from 98.9 percent to 100 percent, NPV ranged from 99 percent to 100 percent, sensitivity ranged from 99 percent to 100 percent, and specificity ranged from 99 percent to 100 percent.
            - For site 19, the second Meditech site, PPV, NPV, sensitivity, and specificity were all 100 percent.
            - For site 20, the third Meditech site, PPV ranged from 96 to 100 percent, NPV ranged from 99.4 percent to 100 percent, sensitivity ranged from 96 percent to 100 percent, and specificity ranged from 99.4 percent to 100 percent.
    - The developers also sampled 30 denominator-excluded encounters for each Meditech and Cerner sites to assess whether excluded cases for EHR data truly met the intent for exclusion.
        - All 20 sites had a 100 percent match rate for denominator exclusions.
- Validity testing conducted at the Accountable Entity Level:
    - The developers conducted known groups validity testing which focuses on the measure's ability to differentiate between groups of measured entities known to differ on their underlying latent construct.
        - For this measure, the developers used three known groups suggested from prior research including:
        - Hospital teaching/academic status
        - Hospital bed size
        - Hospital urban/rural location
        - The developers hypothesized that risk-adjusted AKI rate would be lower at teaching, large-sized, and urban hospitals than at non-teaching, small-sized rural hospitals.
    - The developers found that teaching hospitals' risk-adjusted AKI rates were 27 percent lower on average than non-teaching hospitals. The developer stated that descriptive findings suggest that large hospitals have better resources and staff mix than small hospitals, however, the developer stated that findings are not conclusive due to relatively small number of facilities.
- The Feasibility Scorecard did not indicate any accuracy issues with the data elements.

Exclusions

- The developers used two methods for testing exclusions:
    - The first was using the full denominator data and removing the exclusion criterion one at a time and calculated a marginal and relative impact on the prevalence of the numerator and denominator as well as the observed measure rate.
        - The average relative impact on the denominator, numerator, and measure rate ranged from less than 1 percent to approximately 40 percent, less than 1 percent to more than 75 percent, and negative 2 percent to nearly 40 percent, respectively.
    - The second was using a parallel-form comparison where the developers determined if exclusions from the denominator per the EHR met the clinical intent for exclusion.
        - All 20 sites had a 100 percent match rate for denominator exclusions.

- The developers attested that all exclusions are necessary to reduce the measure's false positive rate and to exclude patients for whom clinical experts agreed that AKI is essentially nonpreventable.

Risk Adjustment

- The developer conducted a statistical risk model with 28 risk coefficients from 13 different risk factors. The risk model was generated in five steps.
    - The first was to randomly divide the denominator data into a training set, a validation set, and a holdout set.
    - Then the developers ran a locally weighted logistic regression relating measure outcome to each risk factor and examined functional forms.
    - The developers then fit the model using the least absolute shrinkage and selection operate on the training set and examined model fit on the validation set. The validation set also facilitated parameter tuning and the developers considered three common approaches to tuning: 1) minimum of CV function, 2) 'One-Standard-Error' rule, and 3) minimum of Bayesian Information Criterion function.
    - The fourth step was to compare selected risk factors across the three methods, assessed each risk factor's directional impact on the outcome in the training set via a logistic regression, and brought in clinical expertise to finalize factors for the baseline model.
    - The last step was to estimate a final, unbiased assessment of model performance on the holdout set.
- The C-statistic for the model was 0.863 which the developer stated is larger than the benchmark, which is 0.8, frequently cited for demonstrating excellent model performance. The developer also stated that the predictions of the model do not significantly deviate from the observed rate, therefore, the model's external validation is satisfactory.
    - The developers note that because the sample size is relatively small, they performed a simulation exercise to gauge how model calibration may be driven by chance and found that the odds of a strong model performance occurring by chance are very small.
- The developer stated that for risk factors they looked at patient characteristics and mediating factors.
    - For patient characteristics, they chose sex, age, vital signs, index eGFR, and comorbidities at the start of care based on published literature summarized by KDIGO and available data elements in the EHR.
        - The developers did not include race, ethnicity, income, and living environment. The developers stated that for race and ethnicity, there is no theoretical reason to believe African Americans are more likely to experience AKI than Whites and because race was removed from the eGFR, it would be inconsistent to adjust for race. For living environment, the developer stated that it should within hospitals' control to practice best care and adjust care processes for those who are affected by living environment.
        - The developer further stated that the empirical findings suggest that the model does not perform better with these factors included and therefore, inclusion is not justified.
    - For mediating factors, the developers chose hospital length of stay (LOS) as longer LOS can be reflective of the patient's health status not captured by comorbidities or vital signs and ineffective hospital care practices can extend hospital days beyond what was expected or normative.
    - Based on the conceptual model testing, the developer concluded the following regarding the proposed risk factors:
        - Included sex-by-age groups

- All vital signs (in the encounter) except for oxygen saturation were included
- Included index eGFR values
- Included cancer (llm), diabetes, hypertension, obesity, and heart failure
- Excluded AIDS, autoimmune conditions, cancer (other), dementia, drug abuse, chronic pulmonary disease, peripheral vascular disease, and thyroid disorders
- Included hospital LOS

Meaningful Differences

- The developer reported full denominator data was used to calculate hospital-level measure rates and the 95 percent confidence intervals for each of the 20 test sites. The developer then calculated a system-wide weighted average measure rate across the test sites and compared each hospital's performance against the system-wide weighted average and determined if the difference was statistically significant.
- The developer also ran two linear regressions relating incidence of AKI to hospital-specific fixed effects and estimated a generalized T-test. The first regression did not account for patient characteristics while the second regression did.
- The developer reported that testing data show measure performance rates ranged from 0.76 percent to 4.43 percent.
- The developer noted that several hospitals' rates were consistently below or above the weighted mean which the developer attested shows room for improvement in the inpatient setting.
- The developer stated the regression results confirms that the measure can identify clinically meaningful differences across hospitals.

Missing Data

- The developer conducted a parallel-form comparison by comparing data exported from the EHR to data manually abstracted from patients' medical charts. They tabulated the frequency and percentage of data missingness for each key data element.
    - For the 17 Cerner sites, all data elements had zero percent missing data except for one, kidney dialysis ordered or performed DateTime, which had six percent missing data.
    - For site 18, the first Meditech site, all data elements had zero percent missing data except for two, patient eGFR value and patient eGFR DateTime, which both had 3.8 percent missing data.
    - For site 19 and site 20, the second and third Meditech sites, all data elements had zero percent missing data.
- The developer stated that for the site that had missing eGFR values, the site resolved this problem by applying a revised (race-neutral) CKD-EPI formula to serum creatinine values, because they were almost never missing, to calculate eGFR consistently across all observations at all sites. The developer attests that this resolution eliminated any bias due to missing eGFR values.
- The developer stated that for the sites that had missing kidney dialysis ordered or performed was attributed to the use of third-party contracted services to perform dialysis which led to documentation of dialysis only in unstructured form. The developer offers a solution to this problem by stating that facilities should ensure that the performance of dialysis is documented in structured fields.

Comparability

- The measure only uses one set of specifications for this measure.

Questions for the Standing Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?*

Guidance From the Validity Algorithm

All threats to validity were empirically assessed (Box 1) → Empirical validity testing conducts on the measure as specified (Box 2) → Empirical testing conducted at the accountable entity level (Box 5) → Method used to test validity was described and appropriate (Box 6) → Moderate certainty or confidence that the accountable entity levels are a valid indicator of quality (Box 7a) → Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications, including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer stated that the data elements are generated or collected and used by healthcare personnel during the provision of care. Further, the data elements are coded by someone other than the person obtaining original information.
- The developer stated that all data elements are in defined fields in EHRs.
- Using a simulated data set, the submission demonstrates that the evaluation of 100 percent of the measure logic can be automated.
- The Feasibility Scorecard assesses each data element across the following domains:
    - Availability – is the data element readily available in a structured format across electronic health record (EHR) systems?
    - Accuracy – is the information contained in the data correct?
    - Standards – is the data element coded using a nationally accepted terminology standard?
    - Workflow – is the data element routinely captured and used during care delivery?
- The developer identified feasibility issues for the following data elements.
    - Diagnosis data element (ICD-10-CM diagnoses used in measure exclusions) – These data elements are available the EHR but the developer found that the 'Present on Admission' (POA) indication at some sites was not available for extraction during testing. The developer identified two main reasons for this issue but did note that they have since been resolved:
        - The first issue was that the interface allowing the POA indicator to flow into the EHR was not working and required troubleshooting beyond the testing timeframe.
        - The second issue was the lack of access to the POA indicator data source location during testing. The developer noted that they do not have concerns about the availability, accuracy, and use of standards for the POA indicator as it is a required element in hospital billing.
    - Dialysis services not offered at all prospective test sites – The developer states that seven of the 29 sites that offered dialysis services outsourced the services. Because of this, the clinical documentation was not available in structured data. The developer stated there is no concern with this issue since the measure is able to capture the intended dialysis population through

with this issue since the measure is able to capture the intended dialysis population through ICD-10-CM diagnosis codes.

Questions for the Standing Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form (e.g., EHR or other electronic sources)?*
- *Is the data collection strategy ready to be put into operational use?*
- *For data elements assessed to have feasibility issues, does the developer present a credible, near-term path to electronic collection?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

- Moderate – all identified feasibility issues have a core plan to address the issues and 100 percent coverage in simulated data unit tests (BONNIE)

Criterion 4: Use and Usability

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Current uses of the measure

Publicly reported?	☐	Yes	☒	No		
Current use in an accountability program?	☐	Yes	☒	No	☐	UNCLEAR
Planned use in an accountability program?	☒	Yes	☐	No	☐	NA

Accountability program details

- The developer stated that the measure is not currently used in an accountability program as the measure is under initial endorsement review, however, the measure was submitted to the 2022 Measures Under Consideration (MUC) list and will be reviewed by the Measure Applications Partnership (MAP) during the 2022-2023 review cycle. The developer states that CMS has sought MAP support for implementation in accountability programs such as Hospital Inpatient Quality Reporting and Promoting Interoperability Programs.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: (1) Those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; (2) Those being measured, and other users have been given an opportunity to provide feedback on the measure performance or implementation; and (3) This feedback has been considered when changes are incorporated into the measure.

Feedback on the measure provided by those being measured or others

- The developer noted that CMS collects feedback via Measures Management System (MMS) posting with announcements to stakeholders, NQF endorsement review, MAP review, Proposed Rules published in the Federal Register, QualityNet portal, and a technical advisory panel representing key stakeholders and clinical experts.
- The developer stated that implementation resources are provided through CMS eCQI Resource Center and The ONC Project Tracking System. The developer stated that feedback is evaluated and presented during the CMS Annual Update change review process for measure refinements.
- The developer stated that during development, a technical expert panel composed of a variety of stakeholders was engaged as well as subject matter experts at the Renal Physicians Association and American Society of Nephrology. The developer noted that they also collected feedback from pilot test sites and conducted a survey to inquire about the measure's usability and its prospect of field implementation.
- The developer noted that the feedback received during development was used to make modifications to incorporate more accurate diagnoses of AKI using clinical data, exclusions of pre-hospitalization acquired AKI, and the risk adjustment methodology.

***Questions for the Standing Committee**:*

- *How have (or can) the performance results be used to further the goal of high quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4b1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer stated that because this is a new measure, no trend data is available.

4b2. Benefits versus harms. The benefits of the performance measure in facilitating progress toward achieving high quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer attested that they did not identify any unintended consequences during eCQM development or testing.

Potential harms

- The developer attested that no unexpected benefits were not during development testing.

Additional Feedback:

- The measure was submitted to the MUC list for 2022-2023 MAP review cycle to be considered for the Hospital Inpatient Quality Reporting and Promoting Interoperability programs.

***Questions for the Standing Committee**:*

- *How can the performance results be used to further the goal of high quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for Usability and Use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 5: Related and Competing Measures

Related Measures

- The developer did not note any NQF-endorsed related measures. However, they did note a non-NQF endorsed related measure, Patient Safety Indicator (PSI) 10: Postoperative Acute Kidney Injury Requiring Dialysis Rate. The developer noted that PSI 10 is included in the NQF-endorsed composite measure, NQF #0531 *PSI 90 Patient Safety and Adverse Events*.

Harmonization

- The developer stated that harmonization between PSI 10 and this measure are not necessary because there are differences between the measure and PSI 10, particularly the goal of each of the measures and the data sources used for each.
- The developer further stated that harmonization between this measure and NQF #0531 is not necessary because the only overlapping similarity between the measures is PSI 10 which the developer previously stated does not need harmonization as the outcome of the measure is different. The developer continued noting that while the target populations for the measures are similar, the denominators are different.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a.01. Provide a logic model.

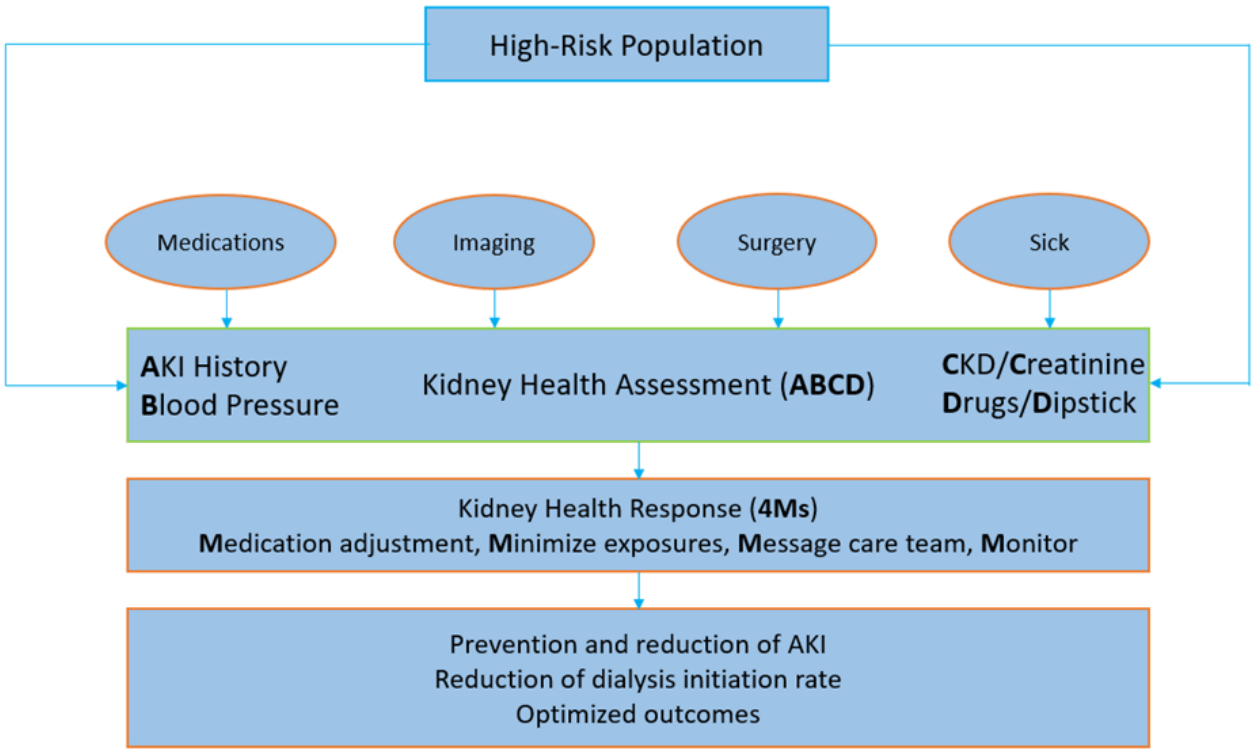
Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]

In the hospital setting, acute kidney injury (AKI) is typically medication induced (nephrotoxic drugs and iodinated contrasts), a result of sepsis following surgery or trauma, or due to hypotension resulting from cardiovascular causes (Singh et al., 2013; Onuigbo et al., 2017). Currently, the lack of quality indicators contributes to considerable variation in care and difficulty in studying what interventions may lead to improved outcomes. Recognizing this gap, the 22nd Acute Disease Quality Initiative meeting was convened in 2018 to discuss the evidence, provide recommendations, and highlight future directions for AKI-related quality measures and care processes. Using a modified Delphi process, an international group of experts including physicians, advanced practice nurses, and pharmacists produced a framework for current and future quality improvement in the area of AKI. Best practices to improve the prevention, identification, and care of patients with AKI were identified and highlighted. These recommendations have been translated into the AKI Logic Model that guided development of this eCQM (adapted from Kashani et al., 2019) to promote prevention of moderate-to-severe AKI in the hospital setting and optimize outcomes for this patient population.

This logic model is based on the concept of regular Kidney Health Assessments in high-risk populations. High-risk populations are defined by exposures such as nephrotoxic medications, imaging with contrast, surgery, and sickness; most hospitalized, non-pregnant adults qualify as high risk based on one or more of these criteria. The KHA includes **A**KI history, **B**lood pressure, **C**hronic kidney disease and serum **C**reatinine level, **D**rug list, and urine **D**ipstick (ABCD). Depending on the results of the KHA, a Kidney Health Response (KHR) is initiated: **m**edication review and adjustment (e.g., withhold nonsteroidal anti-inflammatory drugs if possible), **m**inimize nephrotoxic exposures (e.g., intravenous contrast), **m**essage the health care team to alert them to the high risk of AKI and to optimize the patient's volume status and hemodynamic parameters, and **m**onitor for AKI and its consequences (4Ms). (Kashani, 2019). As described further below, implementation of the 4Ms is expected to lead to primary prevention of AKI (in at least some cases), secondary prevention of progression from stage 1 to stage 2 or higher AKI, reduction of the risk of requiring dialysis initiation, and improved long-term outcomes.

Figure 2: AKI Logic Model



Citation for Figure: Figure adapted from original figure developed and included in Kashani et al., 2019. Acute Disease Quality Initiative 22 www.adqi.org.

References:

1. Kashani, K., Rosner, M. H., Haase, M., Lewington, A., O'Donoghue, D. J., Wilson, F. P., Nadim, M. K., Silver, S. A., Zarbock, A., Ostermann, M., Mehta, R. L., Kane-Gill, S. L., Ding, X., Pickkers, P., Bihorac, A., Siew, E. D., Barreto, E. F., Macedo, E., Kellum, J. A., Palevsky, P. M., … Wu, V. C. (2019). Quality Improvement Goals for Acute Kidney Injury. Clinical journal of the American Society of Nephrology : CJASN, 14(6), 941–953.
2. Onuigbo, M. A., Samuel, E., & Agbasi, N. (2017). Hospital-acquired nephrotoxic exposures in the precipitation of acute kidney injury – A case series analysis and a call for more preventative nephrology practices. Journal of nephropharmacology, 6(2), 90-97.
3. Singh, T. B., Rathore, S. S., Choudhury, T. A., Shukla, V. K., Singh, D. K., & Prakash, J. (2013). Hospital-acquired acute kidney injury in medical, surgical, and intensive care unit: A comparative study. Indian journal of nephrology, 23(1), 24-9.

[Response Ends]

1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

[Response Begins]

In 2019, KDIGO held a conference in Rome, Italy to identify areas where knowledge has significantly advanced since the publication of the 2012 KDIGO Acute Kidney Injury (AKI) Guideline and to outline existing controversies in diagnosis and management of AKI in order to lay the foundation for a future targeted revision of that guideline (KDIGO, 2019a). This

conference was attended by a global panel of multidisciplinary clinical and scientific experts from both academia and healthcare organizations, AKI survivors and caregivers, and also included sponsors from the private sector including healthcare and life science companies (KDIGO, 2019a). This conference focused on key issues relevant to the management of AKI, revisited KDIGO guideline AKI nomenclature and diagnostic criteria, discussed AKI risk stratification and the role of biomarkers in this process, examined the role of resuscitation fluids and nephrotoxins in the critically ill, and addressed timing and modality of kidney replacement therapy in AKI (KDIGO, 2019a; KDIGO, 2019b; Ostermann et al., 2020). The overall findings from the conference suggest that since publication of the KDIGO guidelines in 2012, consensus of clinical opinion and several recent trials support evidence-based fluid management and drug stewardship to reduce the occurrence of AKI.

Additionally, the 22nd Acute Disease Quality Initiative (ADQI) consensus conference met in November 2018 to discuss the evidence, provide recommendations, and highlight future directions for AKI-related quality measures and care processes (Kashani et al., 2019). The conference was attended by a multidisciplinary group of international experts including physicians, basic and clinical scientists, epidemiologists, advanced practice nurses, and pharmacists, who discussed the evidence and used a modified Delphi process to achieve consensus on recommendations for AKI-related quality indicators (QIs) and care processes to improve patient outcomes. The management and secondary prevention of AKI in hospitalized patients were discussed and consensus recommendation statements were provided (Kashani et al., 2019).

References:

1. Kashani, K., Rosner, M. H., Haase, M., Lewington, A., O'Donoghue, D. J., Wilson, F. P., Nadim, M. K., Silver, S. A., Zarbock, A., Ostermann, M., Mehta, R. L., Kane-Gill, S. L., Ding, X., Pickkers, P., Bihorac, A., Siew, E. D., Barreto, E. F., Macedo, E., Kellum, J. A., Palevsky, P. M., … Wu, V. C. (2019). Quality Improvement Goals for Acute Kidney Injury. Clinical journal of the American Society of Nephrology : CJASN, 14(6), 941–953.
2. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group (2012). KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney international, Suppl. 2: 1–141.
3. Kidney Disease: Improving Global Outcomes (KDIGO). (2019a, April).Controversies Conference on Acute Kidney Injury. *Scope of Work*. https://kdigo.org/conferences/aki-conference/
4. Kidney Disease: Improving Global Outcomes (KDIGO) (2019b, April). Controversies Conference on Acute Kidney Injury. *What Has Changed Since the Last AKI Guideline: Web Presentation*. https://kdigo.org/conferences/aki-conference/
5. Ostermann, M., Bellomo, R., Burdmann, E. A., Doi, K., Endre, Z. H., Goldstein, S. L., Kane-Gill, S. L., Liu, K. D., Prowle, J. R., Shaw, A. D., Srisawat, N., Cheung, M., Jadoul, M., Winkelmayer, W. C., Kellum, J. A., & Conference Participants (2020). Controversies in acute kidney injury: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Conference. Kidney international, 98(2), 294–309.

[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]

Published literature suggests that the incidence of AKI is 10-20% in general hospitalized patients and up to 45-50% among critically ill patients (Thongprayoon et al., 2020). The incidence of AKI in hospitalized patients is comparable to the rates of severe sepsis and acute lung injury (McCoy et al., 2010; Hoste & Schurgers., 2008; Chertow et al., 2005; Perzazella, 2012). AKI requiring dialysis and less severe AKI affect approximately 200-300 and 2,000-3,000 per million population per year, respectively (Chertow et al., 2005). Up to two thirds of intensive care patients will develop AKI, which may result in the need for dialysis and is associated with an increased risk of mortality (Hoste & Schurgers., 2008; Wilson et al., 2013). Not all AKI is avoidable, but a substantial proportion of AKI cases are preventable and/or treatable at an early stage to improve outcomes (KDIGO, 2012; Goldstein et al., 2016). Details regarding the proportionate reduction of the incidence of moderate-to-severe AKI with evidence-based processes of care are provided in 1a.01 above.

For example, a 2009 meta-analysis of randomized controlled trials on the effects of perioperative hemodynamic goal-directed therapy for adult surgical patients identified 20 clinical trials involving 4220 patients published between January

1980 and January 2008 (Brienza, et al 2009). In 5 studies, the treatment group received only plasma expanders and/or blood, whereas in 15 studies optimization was obtained with crystalloids, colloids and/or blood, and inotropes with or without vasodilators. The authors found that the odds of postoperative acute renal injury were significantly reduced by perioperative hemodynamic optimization, when compared with control group (OR 0.64; 95% CI, 0.50–0.83; p = 0.0007). No statistical heterogeneity was detected (I^2=0%). Perioperative optimization was effective in reducing renal injury no matter how the outcome was defined, including studies defining renal injury by serum creatinine and/or need of renal replacement therapy only (OR 0.66; 95% CI, 0.50–0.88; p = 0.004).

Since this meta-analysis and the release of the KDIGO 2012 guidelines, new evidence has emerged that has important implications for clinical practice in diagnosing and managing AKI (Ostermann et al., 2020). The effectiveness of the 2012 KDIGO recommendations in preventing AKI was confirmed in small single-center randomized controlled trials (RCTs), such as the Prevention of AKI (PrevAKI) and the Biomarker Guided Intervention to Prevent AKI after Major Surgery (BigpAK) studies (Meersch et al., 2017; Göcze et al., 2018). In the PrevAKI trial involving 276 high-risk adults undergoing cardiac surgery with cardiopulmonary bypass, patients allocated to receive the "KDIGO bundle" (i.e., avoidance of nephrotoxic agents, discontinuation of agents suppressing the renin-angiotensin system for 48 hours after surgery, close monitoring of serum creatinine and urine output, avoidance of hyperglycemia for 72 hours after surgery, consideration of alternatives to radiocontrast agents, close hemodynamic monitoring with optimization of volume status and hemodynamic parameters according to a prespecified algorithm) had significantly lower rates of moderate and severe AKI compared to the control group (29.7% vs 44.9%)(p = 0.009; OR, 0.518 [95% CI, 0.316–0.851]). In the BigpAK trial involving 121 patients with increased risk of AKI after major abdominal surgery, administration of the biomarker-guided KDIGO care bundle significantly reduced the incidence of moderate and severe AKI in the intervention group to 6.7% compared to 19.7% in the standard care group (P = 0.04; OR, 3.43 [95% CI, 1.04–11.32]). In addition, results of RCTs have provided new data relevant to several facets of preventing and managing AKI, including early resuscitation, fluid therapy, prevention of contrast-associated AKI, and timing of acute renal replacement therapy (RRT) (Kellum et al., 2016, Nijssen et al., 2017, Self et al., 2018, Semler et al., 2018, Zarbock et al., 2016, Gaudry et al., 2016, Barbar et al., 2018).

References:

1. Barbar, S. D., Clere-Jehl, R., Bourredjem, A., Hernu, R., Montini, F., Bruyère, R., Lebert, C., Bohé, J., Badie, J., Eraldi, J. P., Rigaud, J. P., Levy, B., Siami, S., Louis, G., Bouadma, L., Constantin, J. M., Mercier, E., Klouche, K., du Cheyron, D., Piton, G., … IDEAL-ICU Trial Investigators and the CRICS TRIGGERSEP Network (2018). Timing of Renal-Replacement Therapy in Patients with Acute Kidney Injury and Sepsis. The New England journal of medicine, 379(15), 1431–1442.
2. Brienza, N., Giglio, M.T., Marucci, M., Fiore, T. (2009). Does perioperative hemodynamic optimization protect renal function in surgical patients? A meta-analytic study. Critical Care Medicine, 37(6), 2079-2090.
3. Chertow, G. M., Burdick, E., Honour, M., Bonventre, J. V., & Bates, D. W. (2005). Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. Journal of the American Society of Nephrology: JASN, 16(11), 3365–3370.
4. Gaudry, S., Hajage, D., Schortgen, F., Martin-Lefevre, L., Pons, B., Boulet, E., Boyer, A., Chevrel, G., Lerolle, N., Carpentier, D., de Prost, N., Lautrette, A., Bretagnol, A., Mayaux, J., Nseir, S., Megarbane, B., Thirion, M., Forel, J., Maizel, J., Yonis, H., Markowicz, P., Thiery, G., Tubach, F., Ricard, J. and Dreyfuss, D., 2016. Initiation Strategies for Renal-Replacement Therapy in the Intensive Care Unit. New England Journal of Medicine, 375(2), pp.122-133.
5. Göcze, I., Jauch, D., Götz, M., Kennedy, P., Jung, B., Zeman, F., Gnewuch, C., Graf, B. M., Gnann, W., Banas, B., Bein, T., Schlitt, H. J., & Bergler, T. (2018). Biomarker-guided Intervention to Prevent Acute Kidney Injury After Major Surgery: The Prospective Randomized BigpAK Study. Annals of surgery, 267(6), 1013–1020.
6. Goldstein, S. L., Mottes, T., Simpson, K., Barclay, C., Muething, S., Haslam, D. B., & Kirkendall, E. S. (2016). A sustained quality improvement program reduces nephrotoxic medication-associated acute kidney injury. Kidney international, 90(1), 212–221.
7. Hoste, E. A., & Schurgers, M. (2008). Epidemiology of acute kidney injury: how big is the problem? Critical care medicine, 36(4 Suppl), S146–S151.
8. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group (2012). KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney international, Suppl. 2: 1–141.
9. Kellum, J. A., Chawla, L. S., Keener, C., Singbartl, K., Palevsky, P. M., Pike, F. L., Yealy, D. M., Huang, D. T., Angus, D. C., & ProCESS and ProGReSS-AKI Investigators (2016). The Effects of Alternative Resuscitation Strategies on

Acute Kidney Injury in Patients with Septic Shock. American journal of respiratory and critical care medicine, 193(3), 281–287.
10. McCoy, A. B., Waitman, L. R., Gadd, C. S., Danciu, I., Smith, J. P., Lewis, J. B., Schildcrout, J. S., & Peterson, J. F. (2010). A computerized provider order entry intervention for medication safety during acute kidney injury: a quality improvement report. American journal of kidney diseases: the official journal of the National Kidney Foundation, 56(5), 832–841.
11. Meersch, M., Schmidt, C., Hoffmeier, A., Van Aken, H., Wempe, C., Gerss, J., & Zarbock, A. (2017). Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. Intensive care medicine, 43(11), 1551–1561.
12. Nijssen, E. C., Rennenberg, R. J., Nelemans, P. J., Essers, B. A., Janssen, M. M., Vermeeren, M. A., Ommen, V. V., & Wildberger, J. E. (2017). Prophylactic hydration to protect renal function from intravascular iodinated contrast material in patients at high risk of contrast-induced nephropathy (AMACING): a prospective, randomised, phase 3, controlled, open-label, non-inferiority trial. Lancet (London, England), 389(10076), 1312–1322.
13. Ostermann, M., Bellomo, R., Burdmann, E. A., Doi, K., Endre, Z. H., Goldstein, S. L., Kane-Gill, S. L., Liu, K. D., Prowle, J. R., Shaw, A. D., Srisawat, N., Cheung, M., Jadoul, M., Winkelmayer, W. C., Kellum, J. A., & Conference Participants (2020). Controversies in acute kidney injury: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Conference. Kidney international, 98(2), 294–309.
14. Perazella M. A. (2012). Drug use and nephrotoxicity in the intensive care unit. Kidney international, 81(12), 1172–1178.
15. Self, W. H., Semler, M. W., Wanderer, J. P., Wang, L., Byrne, D. W., Collins, S. P., Slovis, C. M., Lindsell, C. J., Ehrenfeld, J. M., Siew, E. D., Shaw, A. D., Bernard, G. R., Rice, T. W., & SALT-ED Investigators (2018). Balanced Crystalloids versus Saline in Noncritically Ill Adults. The New England journal of medicine, 378(9), 819–828.
16. Semler, M. W., Self, W. H., Wanderer, J. P., Ehrenfeld, J. M., Wang, L., Byrne, D. W., Stollings, J. L., Kumar, A. B., Hughes, C. G., Hernandez, A., Guillamondegui, O. D., May, A. K., Weavind, L., Casey, J. D., Siew, E. D., Shaw, A. D., Bernard, G. R., Rice, T. W., & SMART Investigators and the Pragmatic Critical Care Research Group (2018). Balanced Crystalloids versus Saline in Critically Ill Adults. The New England journal of medicine, 378(9), 829–839.
17. Thongprayoon, C., Hansrivijit, P., Kovvuru, K., Kanduri, S. R., Torres-Ortiz, A., Acharya, P., Gonzalez-Suarez, M. L., Kaewput, W., Bathini, T., & Cheungpasitporn, W. (2020). Diagnostics, Risk Factors, Treatment and Outcomes of Acute Kidney Injury in a New Paradigm. Journal of clinical medicine, 9(4), 1104.
18. Wilson, F. P., Yang, W., & Feldman, H. I. (2013). Predictors of death and dialysis in severe AKI: the UPHS-AKI cohort. Clinical journal of the American Society of Nephrology: CJASN, 8(4), 527–537.
19. Zarbock, A., Kellum, J. A., Schmidt, C., Van Aken, H., Wempe, C., Pavenstädt, H., Boanta, A., Gerß, J., & Meersch, M. (2016). Effect of Early vs Delayed Initiation of Renal Replacement Therapy on Mortality in Critically Ill Patients With Acute Kidney Injury: The ELAIN Randomized Clinical Trial. JAMA, 315(20), 2190–2199.

[Response Ends]

1b. Gap in Care/Opportunity for Improvement and Disparities

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

The goal of this acute kidney injury (AKI) electronic clinical quality measure (eCQM) is to improve patient safety and prevent patients from developing moderate-to-severe AKI (i.e., stage 2 or greater) during their hospitalization. Early identification and management of at-risk patients is critical, as there is no specific treatment to reverse acute kidney injury once occurred (KDIGO, 2012). The Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guidelines for AKI have cited serum creatinine as an acceptable proxy for defining and monitoring AKI, and have provided detailed clinical guidelines to evaluate and monitor patients at-risk of kidney damage (Lopes & Jorge, 2013; KDIGO, 2012).

This eCQM defines the harm of AKI as patients who have a substantial increase in their serum creatinine level, or have to initiate dialysis within 48 hours or more after the start of the hospitalization. We define a substantial increase in serum

creatinine as a rise of at least at least 2.0 times higher than the lowest prior serum creatinine value, and the increased value is greater than the highest sex-specific normal value for serum creatinine. This eCQM uses a definition of AKI that is consistent with the definition presented in the KDIGO clinical practice guidelines for Stage 2 or greater (KDIGO, 2012). An increase in serum creatinine indicates a reduction in kidney function, sometimes damaging the kidneys so dialysis is required, and is also associated with an increased risk of mortality (KDIGO, 2012). AKI can cause direct patient harm and symptoms associated with volume overload, electrolyte disorders, uremic complications, and drug toxicity (Lopes & Jorge, 2013; KDIGO, 2012; Hoste & De Corte, 2011). AKI has also been associated with longer term harmful outcomes, such as increased odds of death, increased length of hospital stay, and approximately $7,500 in excess hospital costs (Chertow et al., 2005). This eCQM also captures the need to initiate dialysis after 48 hours of hospital care, as one study found that patients who are treated with renal replacement therapy (dialysis) in the intensive care unit (ICU) still have an extremely high mortality rate of 50-60% (Hoste & Schurgers, 2008). The desired outcome of this eCQM is to reduce AKI and dialysis initiation rates.

While not all instances of acute kidney injury (AKI) are avoidable and AKI may be due to the natural progression of underlying illness or a complication of a necessary treatment such as chemotherapy, a proportion of AKI cases are preventable and/or treatable at an early stage to improve outcomes. The 2012 Kidney Disease: Improving Global Outcomes (KDIGO) guidelines suggest careful management of hemodynamic status, fluid balance, and vasoactive medications, along with avoidance of nephrotoxic exposures and drug dose adjustment, for the prevention and early treatment of acute kidney injury (KDIGO, 2012). Specific KDIGO recommendations (level 1 = recommended, level 2 = suggested; A=high quality evidence, B=moderate quality evidence, C=low quality evidence) include:

“3.1.1: In the absence of hemorrhagic shock, we suggest using isotonic crystalloids rather than colloids (albumin or starches) as initial management for expansion of intravascular volume in patients at risk for AKI or with AKI. (2B)

3.1.2: We recommend the use of vasopressors in conjunction with fluids in patients with vasomotor shock with, or at risk for, AKI. (1C)

3.1.3: We suggest using protocol-based management of hemodynamic and oxygenation parameters to prevent development or worsening of AKI in high-risk patients in the perioperative setting (2C) or in patients with septic shock (2C).

3.3.1: In critically ill patients, we suggest insulin therapy targeting plasma glucose 110–149 mg/dl (6.1–8.3 mmol/l). (2C)

3.3.2: We suggest achieving a total energy intake of 20–30 kcal/kg/d in patients with any stage of AKI. (2C)

3.8.1: We suggest not using aminoglycosides for the treatment of infections unless no suitable, less nephrotoxic, therapeutic alternatives are available. (2A)

3.8.2: We suggest that, in patients with normal kidney function in steady state, aminoglycosides are administered as a single dose daily rather than multiple-dose daily treatment regimens. (2B)

3.8.3: We recommend monitoring aminoglycoside drug levels when treatment with multiple daily dosing is used for more than 24 hours. (1A)

3.8.4: We suggest monitoring aminoglycoside drug levels when treatment with single-daily dosing is used for more than 48 hours. (2C)

3.8.5: We suggest using topical or local applications of aminoglycosides (e.g., respiratory aerosols, instilled antibiotic beads), rather than intravenous application, when feasible and suitable. (2B)

3.8.6: We suggest using lipid formulations of amphotericin B rather than conventional formulations of amphotericin B. (2A)

3.8.7: In the treatment of systemic mycoses or parasitic infections, we recommend using azole antifungal agents and/or the echinocandins rather than conventional amphotericin B, if equal therapeutic efficacy can be assumed. (1A)

4.3.2: We recommend using either iso-osmolar or low-osmolar iodinated contrast media, rather than high-osmolar iodinated contrast media in patients at increased risk of CI-AKI. (1B)

4.4.1: We recommend intravenous volume expansion with either isotonic sodium chloride or sodium bicarbonate solutions, rather than no intravenous volume expansion, in patients at increased risk for CI-AKI. (1A)

4.4.2: We recommend not using oral fluids alone in patients at increased risk of CI-AKI. (1C)"

This eCQM focuses on stage 2 or higher AKI to encourage entities to identify high-risk patients, and to diagnose AKI at its earliest stage (stage 1), in order to implement interventions to prevent progression. Accurately monitoring the rate at which AKI occurs in the hospital setting will allow hospitals to improve quality and reduce AKI harm rates. Several studies identified through systematic literature searches developed or evaluated the effectiveness of acute kidney injury electronic alert systems (Selby et al., 2012; Ahmed et al., 2015; Porter et al., 2014; Wilson et al., 2014; Kirkendall et al., 2014; Cho et al., 2012). These studies used data elements for defining AKI that were already present and populated in the EHR. For acute kidney injury diagnosis, all except two were limited to using serum creatinine levels, suggesting that this is the most reliable and consistently available electronic data element for defining acute kidney injury.

References:

1. Ahmed, A., Vairavan, S., Akhoundi, A., Wilson, G., Chiofolo, C., Chbat, N., Cartin-Ceba, R., Li, G., & Kashani, K. (2015). Development and validation of electronic surveillance tool for acute kidney injury: A retrospective analysis. Journal of critical care, 30(5), 988–993.
2. Chertow, G. M., Burdick, E., Honour, M., Bonventre, J. V., & Bates, D. W. (2005). Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. Journal of the American Society of Nephrology: JASN, 16(11), 3365–3370.
3. Cho, A., Lee, J. E., Yoon, J. Y., Jang, H. R., Huh, W., Kim, Y. G., Kim, D. J., & Oh, H. Y. (2012). Effect of an electronic alert on risk of contrast-induced acute kidney injury in hospitalized patients undergoing computed tomography. American journal of kidney diseases: the official journal of the National Kidney Foundation, 60(1), 74–81.
4. Hoste, E. A., & De Corte, W. (2011). Clinical consequences of acute kidney injury. Contributions to Nephrology; 174: 56–64.
5. Hoste, E. A., & Schurgers, M. (2008). Epidemiology of acute kidney injury: how big is the problem? Critical care medicine, 36(4 Suppl), S146–S151.
6. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group (2012). KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney international, Suppl. 2: 1–141.
7. Kirkendall, E. S., Spires, W. L., Mottes, T. A., Schaffzin, J. K., Barclay, C., & Goldstein, S. L. (2014). Development and performance of electronic acute kidney injury triggers to identify pediatric patients at risk for nephrotoxic medication-associated harm. Applied clinical informatics, 5(2), 313–333.
8. Lopes, J. A., & Jorge, S. (2013). The RIFLE and AKIN classifications for acute kidney injury: a critical and comprehensive review. Clinical kidney journal, 6(1), 8–14.
9. Porter, C. J., Juurlink, I., Bisset, L. H., Bavakunji, R., Mehta, R. L., & Devonald, M. A. (2014). A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association, 29(10), 1888–1893.
10. Selby, N. M., Crowley, L., Fluck, R. J., McIntyre, C. W., Monaghan, J., Lawson, N., & Kolhe, N. V. (2012). Use of electronic results reporting to diagnose and monitor AKI in hospitalized patients. Clinical journal of the American Society of Nephrology: CJASN, 7(4), 533–540.
11. Wilson, F. P., Reese, P. P., Shashaty, M. G., Ellenberg, S. S., Gitelman, Y., Bansal, A. D., Urbani, R., Feldman, H. I., & Fuchs, B. (2014). A trial of in-hospital, electronic alerts for acute kidney injury: design and rationale. Clinical trials (London, England), 11(5), 521–529.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

A total of 20 hospitals (test sites) with varying bed size, geographic location, teaching status, urbanicity, and EHR systems participated in measure testing. Using data from test sites' EHR systems over the full calendar year 2020, hospitals' performance rate in AKI ranged from a low (min) of 0.76 to a high (max) of 4.43 per 100 qualified inpatient admissions. The system-wide, weighted average measure rate equaled 1.52 per 100 qualified inpatient admissions. The standard deviation of measure performance rate across sites was 1.01 per 100 qualified inpatient admissions. The interquartile range was 0.66 (unadjusted) and 0.84 (risk-adjusted) per 100 qualified inpatient admissions (i.e., 1.04-1.70 and 1.12-1.96, respectively).

Table 43 below shows high-level characteristics of test sites and their measure performance rate (before risk adjustment and after risk adjustment) based on data from calendar year 2020.

Table 43. High-level Characteristics of Test Sites and Measure Performance Rate (Score) in CY2020

Hospital	Teaching Status	Urban/Rural	Bed Size	No. of Unique Patients	Denominator Count	Observed Measure Rate	Risk-adjusted Measure Rate
1	No	Urban	200-499	3,032	3,535	1.92%	1.95%
2	No	Urban	200-499	6,432	7,420	1.02%	1.14%
3	No	Urban	200-499	3,750	4,269	1.12%	1.08%
4	No	Urban	100-199	3,801	4,365	1.05%	1.32%
5	No	Urban	200-499	2,671	2,985	1.47%	1.48%
6	No	Urban	100-199	1,496	1,712	0.76%	1.06%
7	Yes	Urban	100-199	2,197	2,663	1.46%	1.85%
8	No	Rural	25-99	145	151	3.31%	4.36%
9	Yes	Urban	200-499	3,102	3,727	1.29%	1.03%
10	No	Urban	200-499	4,258	5,060	1.01%	0.89%
11	Yes	Urban	100-199	1,759	2,112	1.37%	1.56%
12	No	Rural	100-199	1,349	1,613	1.05%	1.65%
13	No	Rural	25-99	766	889	1.01%	1.44%
14	Yes	Urban	> 499	6,769	7,948	1.72%	1.13%
15	Yes	Urban	200-499	3,945	4,750	1.60%	1.30%
16	No	Urban	< 25	620	664	1.36%	2.59%
17	No	Rural	100-199	493	551	0.91%	0.99%
18	No	Rural	25-99	517	593	1.69%	1.97%
19	No	Urban	100-199	2,114	2,481	4.43%	3.19%
20	No	Urban	100-199	1,279	1,448	3.87%	2.92%

Note: A total of 20 hospitals with two different EHR systems (Meditech and Cerner) participated in measure testing. Full year of data were used for testing and risk model development.

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

Published literature suggests that the incidence of AKI is 10-20% in general hospitalized patients and up to 45-50% among critically ill patients (Thongprayoon et al., 2020). The incidence of AKI in hospitalized patients is comparable to the rates of severe sepsis and acute lung injury (McCoy et al., 2010; Hoste & Schurgers., 2008; Chertow et al., 2005; Perzazella, 2012). AKI requiring dialysis and less severe AKI affects approximately 200-300 and 2,000-3,000 per million population per year, respectively (Chertow et al., 2005). Up to two thirds of intensive care patients will develop AKI, which may result in the need for dialysis and is associated with an increased risk of mortality (Hoste & Schurgers, 2008; Wilson et al., 2013). Not all AKI is avoidable, but a substantial proportion of AKI cases are preventable and/or treatable at an early stage to improve outcomes (KDIGO, 2012; Goldstein et al., 2016).

At least one prior study of critically ill patients admitted to the intensive care units at six hospitals in four countries applied KDIGO criteria to estimate variation in the incidence of stage 1 or greater AKI (Srisawat et al., 2015). Of the 15,132 critically ill patients in their cohort, 32% developed AKI based on serum creatinine criteria, but this risk varied widely across sites from 14.6 to 43.8%. After adjusting for differences in age, sex, and severity of illness, the odds ratio for AKI continued to vary across centers (odds ratio (OR), 2.57-6.04, $p < 0.001$).

References:

1. Chertow, G. M., Burdick, E., Honour, M., Bonventre, J. V., & Bates, D. W. (2005). Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. Journal of the American Society of Nephrology: JASN, 16(11), 3365–3370.
2. Goldstein, S. L., Mottes, T., Simpson, K., Barclay, C., Muething, S., Haslam, D. B., & Kirkendall, E. S. (2016). A sustained quality improvement program reduces nephrotoxic medication-associated acute kidney injury. Kidney international, 90(1), 212–221.
3. Hoste, E. A., & Schurgers, M. (2008). Epidemiology of acute kidney injury: how big is the problem? Critical care medicine, 36(4 Suppl), S146–S151.
4. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group (2012). KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney international, Suppl. 2: 1–141.
5. McCoy, A. B., Waitman, L. R., Gadd, C. S., Danciu, I., Smith, J. P., Lewis, J. B., Schildcrout, J. S., & Peterson, J. F. (2010). A computerized provider order entry intervention for medication safety during acute kidney injury: a quality improvement report. American journal of kidney diseases: the official journal of the National Kidney Foundation, 56(5), 832–841.
6. Perazella M. A. (2012). Drug use and nephrotoxicity in the intensive care unit. Kidney international, 81(12), 1172–1178.
7. Srisawat N, Sileanu FE, Murugan R, Bellomod R, Calzavacca P, Cartin-Ceba R, Cruz D, Finn J, Hoste EE, Kashani K, Ronco C, Webb S, Kellum JA; Acute Kidney Injury-6 Study Group. Variation in risk and mortality of acute kidney injury in critically ill patients: a multicenter study. Am J Nephrol. 2015;41(1):81-8. doi: 10.1159/000371748.
8. Thongprayoon, C., Hansrivijit, P., Kovvuru, K., Kanduri, S. R., Torres-Ortiz, A., Acharya, P., Gonzalez-Suarez, M. L., Kaewput, W., Bathini, T., & Cheungpasitporn, W. (2020). Diagnostics, Risk Factors, Treatment and Outcomes of Acute Kidney Injury in a New Paradigm. Journal of clinical medicine, 9(4), 1104.
9. Wilson, F. P., Yang, W., & Feldman, H. I. (2013). Predictors of death and dialysis in severe AKI: the UPHS-AKI cohort. Clinical journal of the American Society of Nephrology: CJASN, 8(4), 527–537.

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for

improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

A total of 20 hospitals (test sites) with varying bed size, geographic location, teaching status, urbanicity, and EHR systems participated in measure testing. **Table 44** below provides detailed information about measure denominator population, stratified by sex, age bins, race/ethnicity, and primary source of payment.

Table 44. Measure Denominator Population Characteristics

Measure Denominator Population Characteristics	**EHR System: Meditech**	**EHR System: Meditech**	**EHR System: Cerner**	**EHR System: Cerner**
*****	**n**	**%**	**n**	**%**
Number of denominator encounters	4,522	100%	54,414	100%
Number of unique patients	3,909	100%	45,398	100%
Age Mean (Std.Dev)	62.5 (16.6)	*	60.4 (17.5)	*
Age bins	*	*	*	*
18-35	298	8%	4,880	11%
36-64	1,766	45%	20,420	45%
65+	1,850	47%	20,099	44%
Sex	*	*	*	*
Male	2,008	51%	23,824	52%
Female	1,902	49%	21,574	48%
Race	*	*	*	*
White	2,858	73%	31,359	69%
Black or African American	440	11%	2,884	6%
Other	603	15%	7,646	17%
Unknown	8	0%	3,526	8%
Ethnicity	*	*	*	*
Hispanic or Latino	491	13%	8,545	19%
Non-Hispanic	3,298	84%	32,404	71%
Unknown	120	3%	4,467	9.8%
(Primary) Payer	*	*	*	*
Medicare	899	23%	8,444	19%
Medicaid	617	16%	15,397	34%
Private Insurance	947	24%	14,035	31%
Self-pay or Uninsured	9	0%	7,695	17%
Other	136	3%	1,722	4%

Measure Denominator Population Characteristics	EHR System: Meditech	EHR System: Meditech	EHR System: Cerner	EHR System: Cerner
Unknown	1,302	33%	0	0%

Note: *Cells intentionally left empty.

Across test sites and within the measure denominator population, male patients showed a slightly higher chance of developing hospital acquired (HA) AKI than female patients and patients aged 36 or above were slightly more likely to develop HA-AKI than those 35 or younger. There was not a noticeable difference in the rate of AKI between Whites and African Americans, Hispanics and non-Hispanics, or Medicare beneficiaries and Medicaid beneficiaries. **Table 45** below provides detailed information on measure performance rate, stratified by sex, age bins, race/ethnicity, and primary source of payment.

Table 45. Measure Performance Rate (Score) Overall and Stratified

*	EHR System: Meditech	EHR System: Meditech	EHR System: Cerner	EHR System: Cerner
Rate of AKI per 100 denominator enctrs	**Rate (%)**	**Std.Err**	**Rate (%)**	**Std.Err**
Across denominator encounters	3.89	0.29	1.32	0.05
Sub-groups	*	*	*	*
Age bins	*	*	*	*
18-35	1.52	0.68	1.05	0.14
36-64	3.95	0.43	1.38	0.07
65+	4.21	0.43	1.33	0.07
Sex	*	*	*	*
Male	4.04	0.41	1.42	0.07
Female	3.74	0.40	1.22	0.07
Race	*	*	*	*
White	3.20	0.31	1.27	0.06
Black or African American	5.25	0.97	1.22	0.18
Other	6.25	0.93	1.57	0.13
Unknown	0.00	0.00	1.40	0.19
Ethnicity	*	*	*	*
Hispanic or Latino	5.21	0.96	1.46	0.12
Non-Hispanic	3.65	0.30	1.29	0.06
Unknown	5.60	2.06	1.32	0.16
(Primary) Payer	*	*	*	*
Medicare	4.10	0.61	1.38	0.12
Medicaid	3.83	0.71	1.33	0.08
Private Insurance	3.73	0.57	1.43	0.09
Self-pay or Uninsured	0.00	0.00	1.13	0.12

*	EHR System: Meditech	EHR System: Meditech	EHR System: Cerner	EHR System: Cerner
Other	5.00	1.73	0.88	0.21
Unknown	3.81	0.50	*	*

Note: Std.Err = standard error, *Cells intentionally left empty.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

N/A

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Hospital Harm - Acute Kidney Injury

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

This electronic clinical quality measure (eCQM) assesses the proportion of inpatient hospitalizations for patients 18 years of age or older who have an acute kidney injury (stage 2 or greater) that occurred during the encounter. Acute kidney injury (AKI) stage 2 or greater is defined as a substantial increase in serum creatinine value, or by the initiation of kidney dialysis (continuous renal replacement therapy (CRRT), hemodialysis or peritoneal dialysis).

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Renal: Acute Kidney Injury

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Safety: Complications

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Adults (Age >= 18)

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Facility

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Inpatient/Hospital

[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]

Final measure specifications for implementation will be made publicly available on CMS' appropriate quality reporting website, once finalized through the NQF endorsement and CMS rulemaking processes.

[Response Ends]

sp.10. Indicate whether Health Quality Measure Format (HQMF) specifications are attached.

Attach the zipped output from the eCQM authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications).

[Response Begins]

HQMF specifications are attached.

[Response Ends]

Attachment: 3713e_AKI-v0-1-126-QDM-5-6_To NQF_(1).zip

sp.11. Attach the simulated testing attachment.

All eCQMs require a simulated testing attachment to confirm that the HTML output from Bonnie testing (or testing of some other simulated data set) includes 100% coverage of measured patient population testing, with pass/fail test cases for each sub-population. This can be submitted in the form of a screenshot.

[Response Begins]

Testing is attached

[Response Ends]

Attachment: 3713e_3713e_AKI Bonnie Testing Coverage SBAR_To NQF_(1)-508.docx

Attachment: 3713e_3713e_Bonnie v5.1.1_ Measure View - CMS832v0_To NQF_(1)-508.pdf

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 3713e_3713e_AKI Value Set Directory_To NQF-508.xlsx

Attachment: 3713e_3713e_2022_NQF_ITS_Attachment_To NQF_(5)-508.xlsx

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.13. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]

Inpatient hospitalizations for patients who develop acute kidney injury (AKI) (stage 2 or greater) during the encounter, as evidenced by:

- A subsequent increase in serum creatinine value at least 2 times higher than the lowest serum creatinine value, and the increased value is greater than the highest sex-specific normal value for serum creatinine; or
- Kidney dialysis (continuous renal replacement therapy (CRRT), hemodialysis or peritoneal dialysis) initiated 48 hours or more after the start of the encounter.

[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

This is an eCQM, and therefore uses electronic health record data to calculate the measure score. The time period for data collection is during an inpatient hospitalization, beginning at hospital arrival including time in the emergency department or observation when these encounters are within an hour of the inpatient admission.

All data elements necessary to calculate this numerator are defined within value sets available in the Value Set Authority Center (VSAC) and listed below.

Serum creatinine tests are represented by LOINC Codes in the value set Creatinine Mass Per Volume (2.16.840.1.113762.1.4.1248.21).

Kidney dialysis is defined by the value set Hospital based dialysis services (2.16.840.1.113762.1.4.1179.4).

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/.

[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Inpatient hospitalizations for patients 18 years of age or older at the start of the encounter without a diagnosis of obstetrics, with a length of stay of 48 hours or longer who had at least one serum creatinine value after 48 hours from the start of the encounter.

[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.16. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

This measure includes all inpatient hospitalization for patients aged 18 years and older at the start of the encounter, and all payers. Inpatient hospitalizations include time in the emergency department and observation when the transition between these encounters (if they exist) and the inpatient encounter are within an hour or less of each other.

Measurement period is one year. This measure is at the inpatient encounter level and measure rates are reported at the hospital level.

- Inpatient encounters are represented using the value set of Encounter Inpatient (2.16.840.1.113883.3.666.5.307).
- Emergency department encounters are represented using the value set of Emergency Department Visit (2.16.840.1.113883.3.117.1.7.1.292).
- Observation encounters are represented using the value set of Observation Services (2.16.840.1.113762.1.4.1111.143).
- Obstetric diagnoses are defined by the value set of Obstetrics and VTE Obstetrics (2.16.840.1.113762.1.4.1248.33)
- Serum creatinine tests are represented by LOINC Codes in the value set Creatinine Mass Per Volume (2.16.840.1.113762.1.4.1248.21).

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/.

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

Denominator exclusions are as follows:

- Inpatient hospitalizations for patients with an increase in serum creatinine value of at least 0.3 mg/dL between the index serum creatinine and a subsequent serum creatinine taken within 48 hours of the encounter start (indicating AKI present on admission).
- Inpatient hospitalizations for patients with the index estimated glomerular filtration rate (eGFR) value of <60 mL/min within 48 hours of the encounter start (indicating chronic kidney disease, stage 3a or greater, present on admission).
- Inpatient hospitalizations for patients who have less than two serum creatinine results within 48 hours of the encounter start (indicating that the hospital stay was too short to diagnose AKI).
- Inpatient hospitalizations for patients who have kidney dialysis (CRRT, hemodialysis or peritoneal dialysis) initiated within 48 hours of the encounter start (indicating end stage renal disease, a severe acute metabolic derangement, or AKI present on admission).
- Inpatient hospitalizations for patients with at least one specified diagnosis present on admission that puts them at extremely high risk for AKI:
    - Hemolytic Uremic Syndrome (HUS)
    - Large Body Surface Area (BSA) Burns
    - Traumatic Avulsion of Kidney
    - Rapidly Progressive Nephritic Syndrome
    - Thrombotic Thrombocytopenic Purpura
- Inpatient hospitalizations for patients who have at least one specified procedure during the encounter that puts them at extremely high risk for AKI:
    - Extracorporeal membrane oxygenation (ECMO)
    - Intra-Aortic Balloon Pump
    - Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA)
    - Nephrectomy

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

To qualify for the denominator exclusions: The serum creatinine values, eGFR values, initiation of kidney dialysis, and procedures that put the patient at high risk for AKI must be present or occur during the qualifying inpatient hospitalization. Diagnoses that put patients at high risk for AKI must be present on admission to the qualifying inpatient hospitalization.

- Serum creatinine tests are represented by LOINC Codes in the value set Creatinine Mass Per Volume (2.16.840.1.113762.1.4.1248.21).

All eGFR values are calculated using the following race-neutral equation, published in 2021[1]:

$eGFR = \mu \times min(\frac{SCr}{k}, 1)^{a_1} \times max(\frac{SCr}{k}, 1)^{a_2} \; x \; c^{Age} \; x \; d[if \; female]$ where μ equals 142, SCr denotes the index serum creatinine value, k equals 0.7 for female and 0.9 for male, a^1 equals -0.241 for female and -0.302 for male, a^2 equals -1.2, c equals 0.9938, age denotes patient age at the start of encounter, and d equals 1.012 and 1 for female and male patients, respectively. Function $min(\cdot)$ selects the lesser of the two parameters and $max(\cdot)$ selects the larger of the two. Because eGFRs are formula based, the "index" eGFR is driven by the index serum creatinine.

- Kidney dialysis is defined by the value set Hospital based dialysis services (2.16.840.1.113762.1.4.1179.4).
- Procedures that put the patient at high risk for AKI are defined by the value set High Risk Procedures for AKI (2.16.840.1.113762.1.4.1248.19)
- Diagnoses that put the patient at high risk for AKI are defined by the value set High Risk Diagnosis for AKI (2.16.840.1.113762.1.4.1248.12)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/.

Reference:

1. Inker, L. A., Eneanya, N. D., Coresh, J., Tighiouart, H., Wang, D., Sang, Y., ... & Levey, A. S. (2021). New creatinine- and cystatin C–based equations to estimate GFR without race. *New England Journal of Medicine*, *385*(19), 1737-1749.

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

N/A. This measure is not stratified.

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

Statistical risk model

[Response Ends]

sp.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

Attachment: 3713e_3713e_2022_NQF_ITS_Attachment_To NQF_(7)-508.xlsx

sp.23. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Lower score

[Response Ends]

sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

Target population: Inpatient hospital encounters, all payers, for individuals who are 18 years or older at the start of the encounter, who do not have an obstetric diagnosis or an extremely high-risk diagnosis or procedure, who have neither acute kidney injury at admission nor stage 3a or higher chronic kidney disease, and who stay in the hospital for at least 48 hours. Inpatient hospitalizations include time in the emergency department and observation when the transition between these encounters (if they exist) and the inpatient encounter are within an hour or less of each other. Inpatient encounters whose discharge dates fell in the measurement period (e.g., 1/1/2020 to 12/31/2020) are considered.

To create the denominator:

1. If the inpatient hospitalization was during the measurement period, go to Step 2. If not, do not include in measure population.
2. Determine the patient's age in years. The patient's age is equal to the encounter start date minus the birth date. If the patient is 18 years or older, go to Step 3. If less than 18 years old, do not include in the measure population.
3. Determine if there is a diagnosis of an obstetric condition during the encounter. If there is not an obstetric diagnosis during the encounter, go to Step 4. If there is an obstetric diagnosis during the encounter, do not include in the measure population.
4. Determine if the encounter length-of-stay is at least 48 hours. If the length of stay is 48 hours or longer, go to Step 5. If the length of stay is less than 48 hours, do not include in the measure population.
5. Determine if there is at least one serum creatinine value after 48 hours from the start of the encounter. If there is not at least one serum creatinine value after 48 hours from the start of the encounter, do not include in the measure population.

Apply denominator exclusions to remove encounters from the denominator:

- Remove encounters for patients with an increase in serum creatinine value of at least 0.3 mg/dL between the index serum creatinine and a subsequent serum creatinine taken within 48 hours of the encounter start. These patients meet KDIGO criteria for acute kidney injury at the start of the encounter. The "index" serum creatinine is defined as the lowest serum creatinine within the first 24 hours of encounter start. If there are no serum creatinine values within the first 24 hours, then the index is the first serum creatinine within the first 48 hours of the encounter start.
- Remove encounters for patients with the index eGFR <60 mL/min within 48 hours of the encounter start. Because eGFRs are formula based, the "index" eGFR is driven by the index serum creatinine. These patients meet KDIGO criteria for chronic kidney disease, stage 3a or greater, at the start of the encounter. The index eGFR is calculated using the index serum creatinine defined above, patient sex, and patient age based on the CKD-EPI Creatinine (Chronic Kidney Disease Epidemiology Collaboration) Creatinine Equations: $eGFR = \mu \times min(\frac{SCr}{k}, 1)^{a_1} \times max(\frac{SCr}{k}, 1)^{a_2} \; x \; c^{Age} \; x \; d[if \; female]$ where μ equals 142, SCr denotes the index serum creatinine value, k equals 0.7 for female and 0.9 for male, a^1 equals -0.241 for female and -0.302 for male, a^2 equals -1.2, c equals 0.9938, age denotes patient age at the start of encounter, and d equals 1.012 and

1 for female and male patients, respectively. Function $min(\cdot)$ selects the lesser of the two parameters and $max(\cdot)$ selects the larger of the two. Because eGFRs are formula based, the "index" eGFR is driven by the index serum creatinine.

- Remove encounters for patients who have less than two serum creatinine results within 48 hours of the encounter start (because it is impossible to rule out AKI at the start of the encounter if the patient does not have at least two serum creatinine results in this period).
- Remove encounters for patients who have kidney dialysis (CRRT, hemodialysis or peritoneal dialysis) initiated within 48 hours of the encounter start (because these patients either have end-stage renal disease, a severe acute metabolic derangement, or AKI at admission). "Initiation" of kidney dialysis is defined as documentation that kidney dialysis (CRRT, hemodialysis or peritoneal) was started during the encounter.
- Remove encounters for patients with at least one specified diagnosis present on admission that puts them at extremely high risk for AKI (i.e., where AKI is expected because of the patient's underlying medical condition):
    - Hemolytic Uremic Syndrome (HUS)
    - Large Body Surface Area (BSA) Burns
    - Traumatic Avulsion of Kidney
    - Rapidly Progressive Nephritic Syndrome
    - Thrombotic Thrombocytopenic Purpura
- Remove encounters for patients who have at least one specified procedure during the encounter that puts them at extremely high risk for AKI (i.e., where AKI is expected because of a procedure to resuscitate the patient by sacrificing renal blood flow, or a procedure for which the therapeutic objective involves losing kidney function):
    - Extracorporeal membrane oxygenation (ECMO)
    - Intra-Aortic Balloon Pump
    - Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA)
    - Nephrectomy

To create the numerator:

To create the numerator, for each encounter identify if the patient develops the harm of AKI stage 2 or greater. There are two ways to meet the numerator harm (AKI stage 2 or greater) criteria. AKI (stage 2 or greater) is defined as a substantial increase in serum creatinine value during the encounter, as evidenced by a subsequent increase in value at least 2 times higher than the lowest serum creatinine value, and the increased value is greater than the highest sex-specific normal value for serum creatinine.

The **first** way to meet numerator:

Diagnose AKI:

1. Evaluate if any serum creatinine value obtained between 48 hours after the start of the encounter and either 30 days after the start of the encounter or discharge, whichever is sooner, is at least 1.5 times higher than the lowest value obtained within the prior 7 days. If yes, then:
2. Evaluate if the increased serum creatinine is greater than the highest sex-specific normal value for serum creatinine. If yes, then:

Stage AKI:

1. Evaluate if any serum creatinine value obtained between 48 hours after the start of the encounter and either 30 days after the start of the encounter or discharge, whichever is sooner, is at least 2 times higher than the lowest value (at any prior time) during the encounter. If yes, then:
2. Evaluate if the increased serum creatinine value is greater than the highest sex-specific normal value for serum creatinine. **If yes, then a numerator harm (AKI stage 2 or greater) has been identified.**

The highest normal serum creatinine value for females is currently defined as 1.02 mg/dL. The highest normal serum creatinine value for males is currently defined as 1.18 mg/dL.

The **second** way to meet numerator:

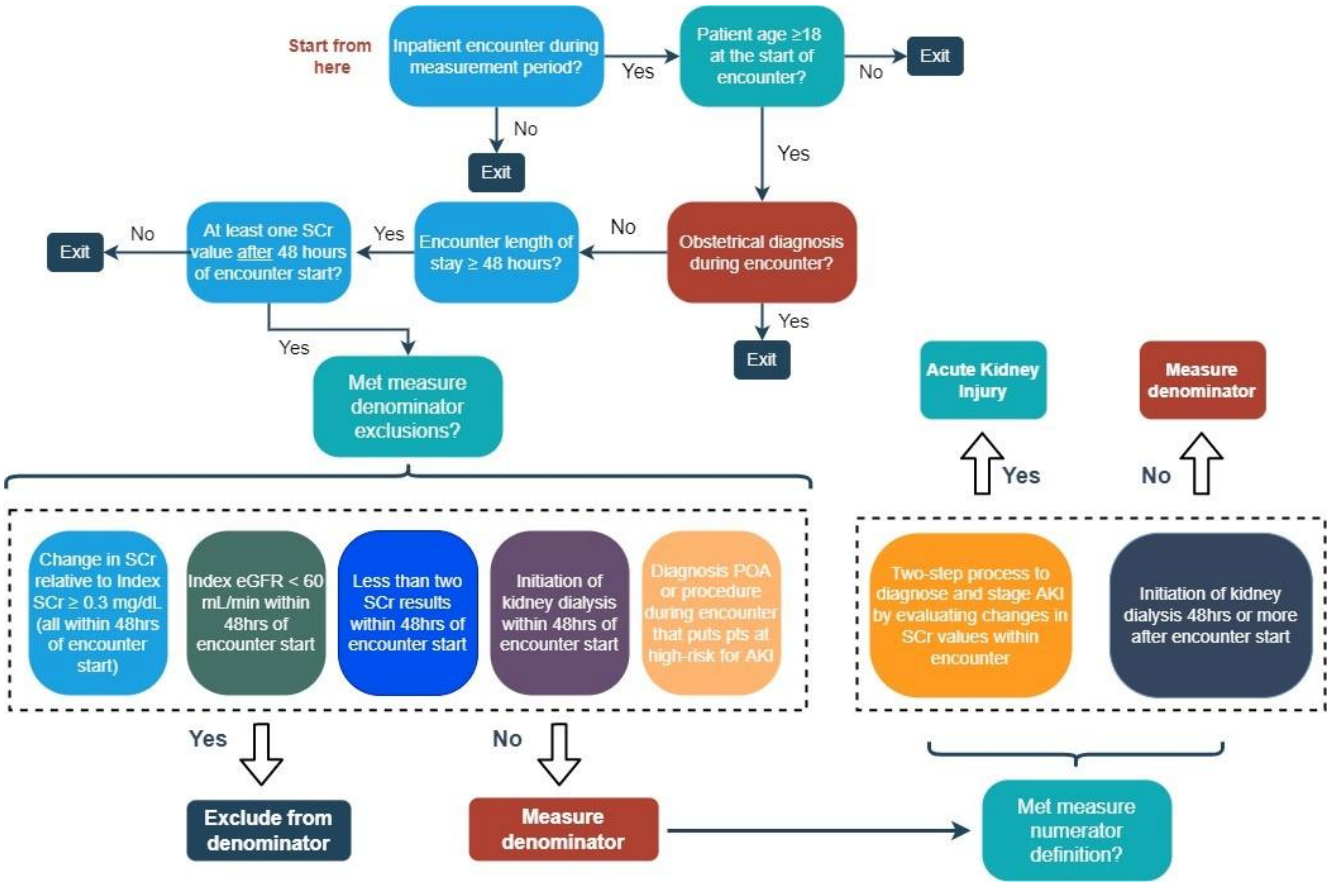
- The initiation of kidney dialysis **48 hours or more after the start of the encounter.**

Only one numerator harm event (AKI) is counted for a given qualifying encounter.

To calculate the hospital-level measure result, divide the total numerator events by the total number of qualifying encounters (denominator).

Please see **Figure 1,** the Hospital Harm-Acute Kidney Injury measure flow diagram below:

Figure 1. AKI Measure Numerator and Denominator Flow Diagram



Note: SCr = serum creatinine; POA = present on admission

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

[Response Begins]

Not applicable.

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Electronic Health Records

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

Hospitals collect EHR data using certified electronic health record technology (CEHRT). The MAT output, which includes the human readable and XML artifacts of the clinical quality language (CQL) for the measure are contained in the eCQM specifications attached. No additional tools are used for data collection for eCQMs.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is

precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face

validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Electronic Health Records

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

Measure development and testing used data from the electronic health records (EHRs) over the full calendar year of 2020 from each of the 20 hospitals (test sites) with two different EHR systems. No hospitals contributed partial year of data. A total of 20 hospitals (test sites) with varying bed size, geographic location, teaching status, urbanicity, and EHR systems participated in measure testing.

Table 1 below provides the high-level information on the data sets for data from calendar year 2020.

Table 1. High-level Characteristics of Data Set from Test Sites

Hospital	Teaching Status	Urban/Rural	Bed Size	Num. of Unique Patients	Num. of Denominator Eligible Encounters
1	No	Urban	200-499	3,032	3,535

Hospital	Teaching Status	Urban/Rural	Bed Size	Num. of Unique Patients	Num. of Denominator Eligible Encounters
2	No	Urban	200-499	6,432	7,420
3	No	Urban	200-499	3,750	4,269
4	No	Urban	100-199	3,801	4,365
5	No	Urban	200-499	2,671	2,985
6	No	Urban	100-199	1,496	1,712
7	Yes	Urban	100-199	2,197	2,663
8	No	Rural	25-99	145	151
9	Yes	Urban	200-499	3,102	3,727
10	No	Urban	200-499	4,258	5,060
11	Yes	Urban	100-199	1,759	2,112
12	No	Rural	100-199	1,349	1,613
13	No	Rural	25-99	766	889
14	Yes	Urban	> 499	6,769	7,948
15	Yes	Urban	200-499	3,945	4,750
16	No	Urban	< 25	620	664
17	No	Rural	100-199	493	551
18	No	Rural	25-99	517	593
19	No	Urban	100-199	2,114	2,481
20	No	Urban	100-199	1,279	1,448

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

01-01-2020 – 12-31-2020

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Facility

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

A total of 20 hospitals (test sites) with varying bed size, geographic location, teaching status, urbanicity, and EHR systems participated in measure testing. Using data from test sites' EHR systems over the full calendar year 2020. A majority of test sites were in the Western U.S. Of the 20 test sites, 15 are in urban areas and five are teaching hospitals. Bed sizes ranged from a low of less than 25 beds to a high of over 499 beds.

- Testing data came from test sites' EHR systems. Testing data in full calendar year 2020 (Jan 1, 2020 to Dec 31, 2020) were used. No partial year data were used.
- A total of 20 hospitals participated in measure testing.
- The number of **unique** patients included in measure denominator ranged from a low of 145 to a high of 6,769 across sites.
- Measure denominator encounters ranged from a low of 151 to a high of 7,948 across sites.

Table 2 below shows high-level characteristics of test sites based on data from calendar year 2020.

Table 2. High-level Characteristics of Test Sites

Test Site	EHR System	Census Region	Bed Size	Teaching Status	Urban/Rural
1	Cerner	West	200-499	Non-academic	Urban
2	Cerner	West	200-499	Non-academic	Urban
3	Cerner	West	200-499	Non-academic	Urban
4	Cerner	West	100-199	Non-academic	Urban
5	Cerner	West	200-499	Non-academic	Urban
6	Cerner	West	100-199	Non-academic	Urban
7	Cerner	West	100-199	Academic	Urban
8	Cerner	West	25-99	Non-academic	Rural
9	Cerner	West	200-499	Academic	Urban
10	Cerner	West	200-499	Non-academic	Urban
11	Cerner	West	100-199	Academic	Urban
12	Cerner	West	100-199	Non-academic	Rural
13	Cerner	West	25-99	Non-academic	Rural
14	Cerner	West	> 499	Academic	Urban

Test Site	EHR System	Census Region	Bed Size	Teaching Status	Urban/Rural
15	Cerner	West	200-499	Academic	Urban
16	Cerner	West	< 25	Non-academic	Urban
17	Cerner	West	100-199	Non-academic	Rural
18	Meditech	Midwest	25-99	Non-academic	Rural
19	Meditech	Northeast	100-199	Non-academic	Urban
20	Meditech	South	100-199	Non-academic	Urban

Note: A total of 20 hospitals with two different EHR systems (Meditech and Cerner) participated in measure testing. Full year of data were used for testing and risk model development.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

Across the 20 test sites and within the measure denominator population, the average patient age was 61 (standard deviation 17). Slightly over half (53%) of the denominator patients were male and approximately 60% of the denominator patients were non-Hispanic Whites. We note that measure testing used a full year of data from each test site, capturing the universe of inpatient discharges in that year.

Table 3. Measure Denominator Population Characteristics

Measure Denominator Population Characteristics	EHR System: Meditech	EHR System: Meditech	EHR System: Cerner	EHR System: Cerner
*	n	%	n	%
Number of denominator encounters	4,522	100%	54,414	100%
Number of unique patients	3,909	100%	45,398	100%
Age Mean (Std. Dev)	62.5 (16.6)	*	60.4 (17.5)	*
Age bins	*	*	*	*
18-35	298	8%	4,880	11%
36-64	1,766	45%	20,420	45%
65+	1,850	47%	20,099	44%
Sex	*	*		
Male	2,008	51%	23,824	52%
Female	1,902	49%	21,574	48%
Race	*	*	*	*

Measure Denominator Population Characteristics	EHR System: Meditech	EHR System: Meditech	EHR System: Cerner	EHR System: Cerner
White	2,858	73%	31,359	69%
Black or African American	440	11%	2,884	6%
Other	603	15%	7,646	17%
Unknown	8	0%	3,526	8%
Ethnicity	*	*	*	*
Hispanic or Latino	491	13%	8,545	19%
Non-Hispanic	3,298	84%	32,404	71%
Unknown	120	3%	4,467	9.8%
(Primary) Payer	*	*	*	*
Medicare	899	23%	8,444	19%
Medicaid	617	16%	15,397	34%
Private Insurance	947	24%	14,035	31%
Self-pay or Uninsured	9	0%	7,695	17%
Other	136	3%	1,722	4%
Unknown	1,302	33%	0	0%

Notes: std.dev = standard deviation. Not all bins total to 100% due to rounding. *Cells intentionally left empty.

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

Measure score level reliability and validity testing used data from the full denominator population. Measure data element level validity testing was based on subsamples drawn from the measure initial population using the approach of random sampling without replacement. These subsamples served as the foundation upon which clinical abstractors compared data exported from the EHR (eData) to data manually abstracted from patients' medical charts (mData, or "gold standard"). This latter process is commonly referred to as the parallel-form comparison.

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

We collected patient race, ethnicity, and primary source of payment from structured fields in the electronic health record. We also identified selected social determinants (e.g., homelessness) that were reported using ICD-10-CM Z codes (Z59). Patient-reported data were not available.

In the measure risk model evaluation, we examined the marginal impact on model performance of adding patient race, ethnicity, primary source of payment, and variables reflecting their social determinants of health (ICD-10-CM Z59). We discuss these findings in the risk adjustment section (2b.32).

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

Data element reliability testing is not necessary for eCQMs that use fully structured data in the EHR. Please see sections 2b.02 and 2b.03 for data element validity testing.

To assess the measure score level reliability, we used Adams' signal-to-noise ratio (SNR) and the intra-class correlation coefficient (ICC) via the split-half sample approach.

We provide reliability statistics for both the unadjusted measure rates and the risk-adjusted rates given that the baseline risk model we developed is not considered final (see section 2a.11). We used the ICC via split-half approach to assess the score-level reliability for the risk-adjusted rates.

To implement an empirical approach based on Adams' SNR, consider that each hospital has a true measure performance rate p that follows a beta distribution. The true rate varies from hospital to hospital due to variation in hospital quality of care in general or variation in the extent to which hospitals exert efforts to prevent AKI in particular. The observed measure rate $\hat{p}$, on the other hand, is binomially distributed (whether or not AKI occurred) conditional on the true rate p. Observed rate $\hat{p}$ also varies in any given time period (e.g., calendar year 2020) due to small number of events occurring in a single facility and random variation around the true rate p.

From the above setup, the alpha and beta parameters underlying the beta distribution can be estimated and then used to calculate the hospital-to-hospital variance, which is frequently known as the signal. This signal records the proportion of variability across measured entities that is attributable to the real difference in quality of care. The hospital-specific, or within-hospital, variance (or noise) can be calculated from the conventional method for any random binomial variable. Therefore, $o^2 hospital - to - hospital = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$ where α and β are the estimated alpha and beta parameters within the testing data, and $\sigma^2 within - hospital = \frac{p\hat{}(1-p\hat{})}{n}$ where $\hat{p}$ is the observed measure rate for a

given hospital and n is the denominator size for that hospital. Reliability, or SNR, is thus equal to $\frac{\sigma^2 hospital-to-hospital}{\sigma^2\ hospital-to-hospital+\ \sigma^2 within-hospital)}$.

To estimate the empirical ICC based on the split-half sample approach, we consider that hospital $h_i(i = 1, \dots, H)$ in subsample $T_i(i = 1, \dots, T)$ and each hospital subsample T_i is comprised of a possibly varying number of denominator encounters n_{ht}. We assume that the measure performance rate, y_{ht}, follows a simple two-level model: $Y_{ht} = \mu + \alpha_h + \varepsilon_{ht}$ where the hospital-level effects a_h are sampled from a normal distribution with mean 0 and variance σ_h^2 and the residual errors are independently and normally distributed with mean 0 and variance $\frac{(\sigma_e^2)}{n_{ht}{}^1}$.

The subsamples here could come from different calendar periods or from randomly generated subsamples (e.g. split-halves) of all denominator encounters, stratified by hospital. Note that the specification of residual error variance assumes that, conditional on the hospital random effects a_h, the variance is *inversely* proportional to the sample size used to form the hospital-subsample estimate. Although such a model can be directly calculated by assuming that encounter-level data follow the standard two-level model for normally distributed data (frequently used in classical testing theory), and that encounter-level data from the same hospital and subsamples are then averaged to form the estimated hospital performance, the proposed model can apply more generally.

The two variance components σ_h^2 and σ_e^2 can be estimated by any statistical software that is capable of fitting maximum likelihood methods. By deriving the estimates of σ_h^2 and σ_e^2, we then compute a "plug-in" estimator of the ICC for performance indicator $CC_h = \frac{\sigma_h^2}{\sigma_h^2+\frac{\sigma_e^2}{n}=\frac{nR}{nR+1}}$, where $R = \frac{\sigma_h^2}{\sigma_e^2}$. Note that ICC is a function only of the size of the denominator and the ratio of between-hospital to within-hospital variance.

When calculating ICC using the risk-adjusted measure rates, we turned to the test-retest reliability. Specifically, for every test site, we first created a holdout sample accounting for 50% of the full sample, estimated risk coefficients from the other half sample, and then calculated the risk-adjusted measure rates for every hospital in both samples. Next, we defined noise variance (σ_e^2) as the conditional variance of a measure (in this case, the observed hospital-level risk-adjusted rate) given the true risk-adjusted hospital rate, where the conditional variance is due to sampling error within each hospital. We defined signal variance (σ_h^2) as the between-hospital variance in the true risk-adjusted measure rates (i.e., variation due to hospital performance). We calculated noise variance as the sampling variance of risk-adjusted rates, assuming that each denominator encounter has the probability of suffering HA-AKI estimated from the population risk-adjustment model. We followed an analogous approach to calculating the signal variance and assumed an implicit two-stage model:

- Stage 1: true risk-adjusted hospital rates are approximately normally distributed
- Stage 2: sampled risk-adjusted hospital rates, conditional on the true risk-adjusted hospital rates, are approximately normally distributed

The split-half sample approach (for both the unadjusted measure rates and the risk-adjusted measure rates) results in an ICC formula that is a function of the ratio of estimated variance components and a given hospital's subsample size. Thus, we used the Spearman-Brown prophecy formula and estimated reliability as a function of the hospital subsample size.[2] We applied this methodology to hospital subsamples created by randomly dividing the full year of data (January 1, 2020 – December 31, 2020) for each hospital into two halves. The resulting estimated variance components were based on "six-months" worth of data for a collection of hospitals with varying denominator size. We performed the above steps 100 times to avoid any one-time estimate (either high or low) being driven by chance. We report the average, median, and range of estimated ICC estimates from this simulation process below.

The higher the SNR or ICC the higher the statistical reliability of the measure, and thus the greater the amount of variation is attributable to systematic differences in performance across hospitals (i.e., signal versus noise).

We used the rubric established by Landis and Koch (1977), while acknowledging its limitations, to interpret estimated SNRs and ICCs:[3]

- 0 – 0.2: slight agreement
- 0.21 – 0.39: fair agreement
- 0.4 – 0.59: moderate agreement
- 0.6 – 0.79: substantial agreement

- 0.8 – 0.99: almost perfect agreement
- 1: perfect agreement

References:

1. Dickens, William T. "Error components in grouped data: is it ever worth weighting?." *The Review of Economics and Statistics* (1990): 328-333.
2. “Spearman-Brown Prophecy Formula” in: Frey, B. (2018). The SAGE encyclopedia of educational research, measurement, and evaluation (Vols. 1-4). Thousand Oaks,, CA: SAGE Publications, Inc. doi: 10.4135/9781506326139
3. Landis, J. Richard, and Gary G. Koch. "The measurement of observer agreement for categorical data." *biometrics* (1977): 159-174.

[Response Ends]

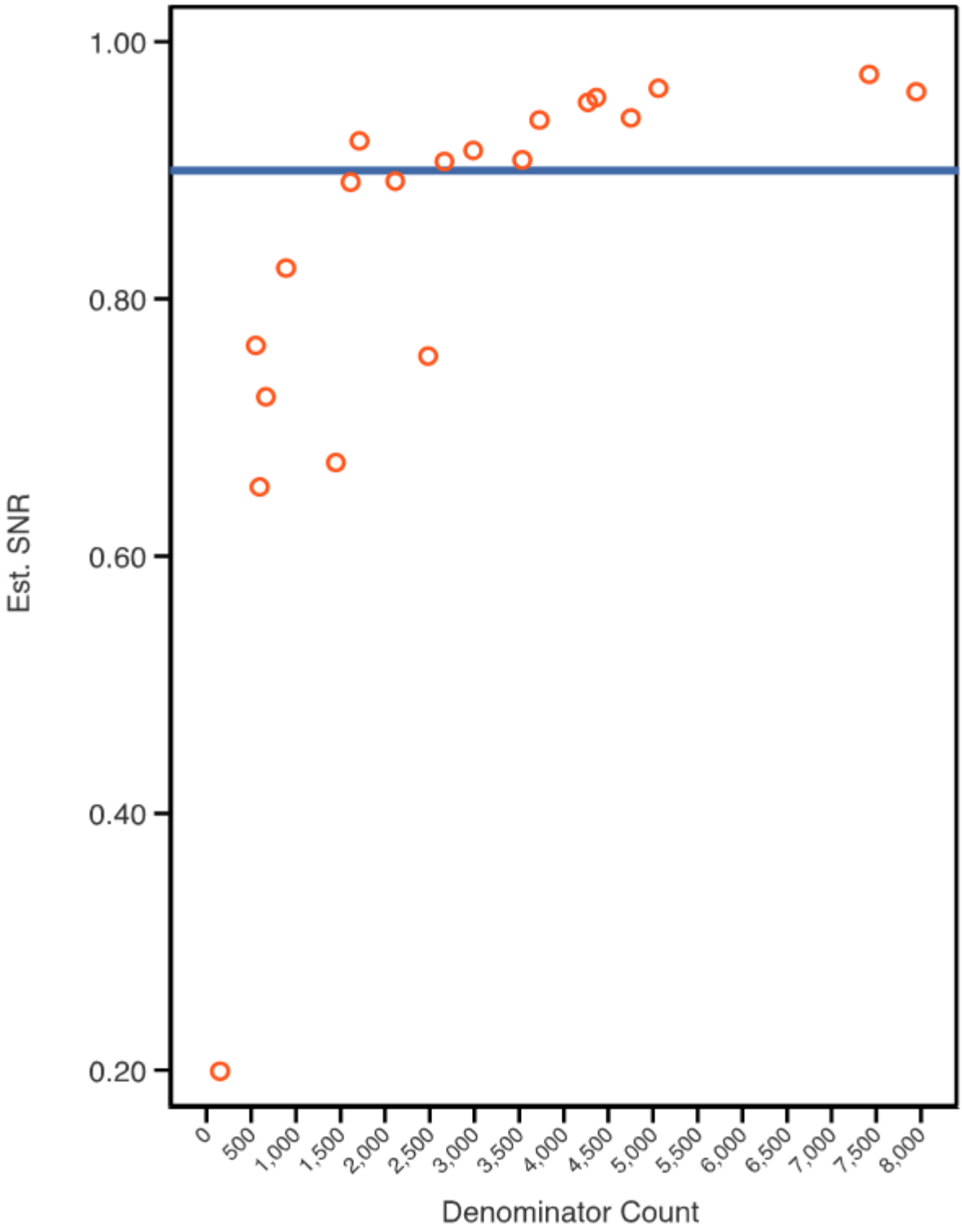
2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

Using the encounter-level data from 20 test sites and in the full calendar year 2020, Adams’ SNRs ranged from 0.20 to 0.97, with the mean and median equal to 0.84 and 0.91 respectively. **Exhibit 1** below shows the distribution of SNRs across test sites with different denominator sizes. The lowest estimated SNR corresponds to a rural site with approximately 150 denominator encounters.

Exhibit 1. Distribution of SNRs Across 20 Hospital Sites



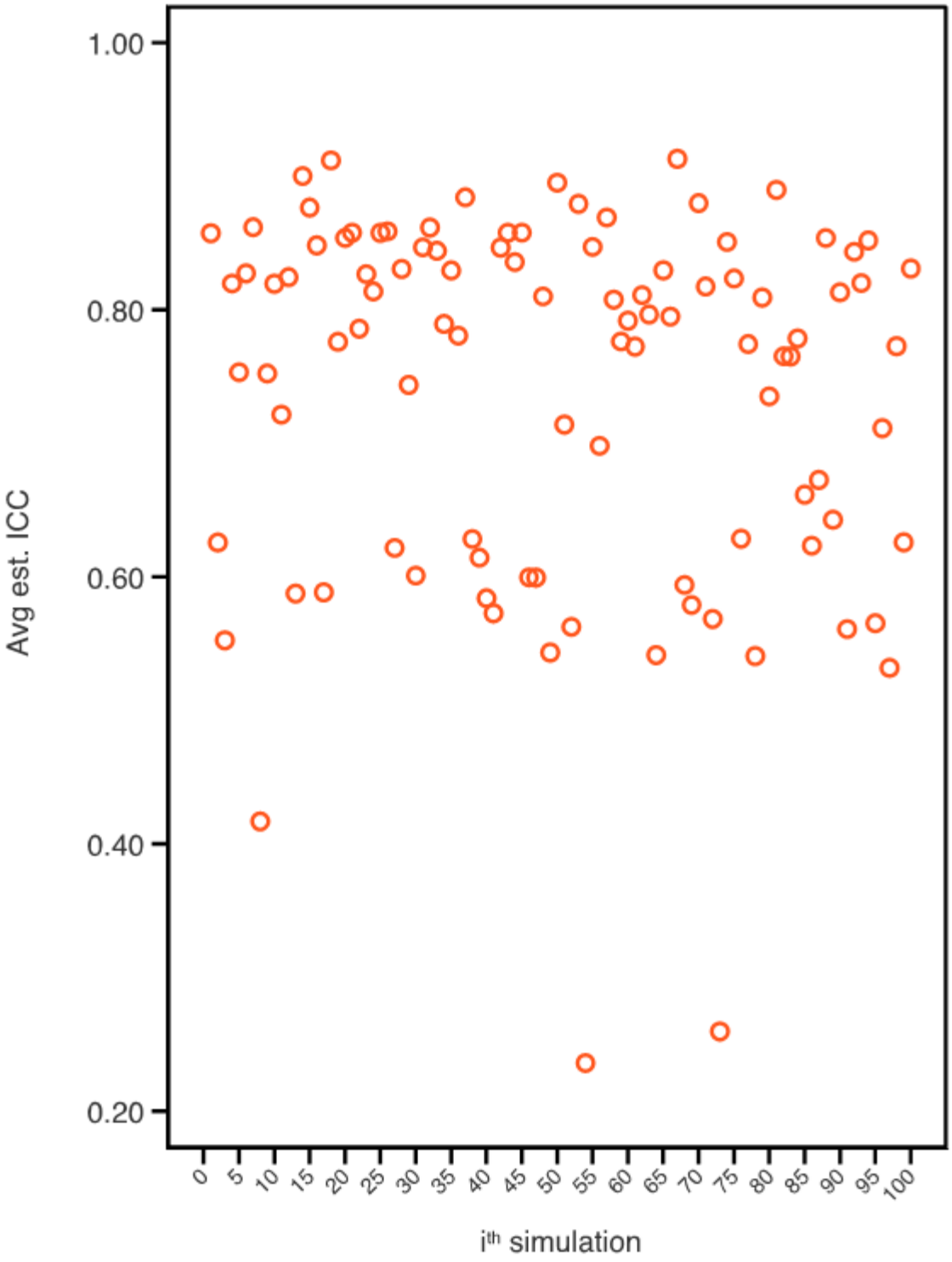
Note: Each circle indicates an estimated SNR. The blue horizontal line denotes the median value of SNR from the distribution.

Empirical findings are supportive, but interpretation requires caution. First, SNR quantifies score-level (i.e., hospital-level) reliability and hence estimation accuracy depends upon the number of hospitals. With only 20 hospitals in testing, the alpha and beta parameters underlying the SNR are estimated with noise. Second, measure performance rates across sites are low but exhibit wide dispersion. To gauge the impact of hospital counts on SNR estimation, we ran two simulation tests. In the first test, we randomly selected a subset of hospitals and estimated each hospital's SNR in that

subsample. We used random sampling with replacement and hence, "small" hospitals (with less than 1,000 denominator encounters) could be selected in or selected out. In our testing data, these are hospitals 8, 13, 16, 17, and 18. The second test was similar except that we always included "small" hospitals in the subsample. The number of hospitals included ranged from six to the full set of 20. Testing results showed that the median SNR is virtually always above 0.8 (**2022 NQF ITS Attachment: Sheet 5)**.

Next, we estimated ICCs for observed measure rates using the split-half sample approach, repeating the procedure 100 times. Across these 100 simulations, the median value of the median ICC for observed measure rates approximated 1.0; no simulations generated median ICCs below 0.99. Mean ICC values showed more variation, as expected, ranging from 0.25 to 0.91 across simulations. **Exhibit 2** shows the distribution of mean values of estimated ICC from the 100 simulation runs.

Exhibit 2. Distribution of Average Estimated ICC Via the Split-half Sample Approach



Note: Each circle indicates the average value of ICC across 20 test sites from a given sample split. A total number of 100 different sample splits was performed.

Finally, we calculated ICC based on the risk-adjusted measure rate for each half of the split sample. We used data from one half (randomly selected) of the full sample to estimate coefficients and then calculated risk-adjusted measure rates for both halves (the current half and the remaining half). This process generated two risk-adjusted measure rates for every test site, based on mutually exclusive sets of encounters from the same time period. As above, we performed 100 simulation runs to avoid any single estimate being driven by chance. Across these 100 simulations, the median value of

the median ICC for risk-adjusted measure rates was 0.99; all but two simulations generated median ICC values over 0.95. Mean ICC values showed more variation, as expected, but the median value (across 100 simulations) of the mean ICC for risk-adjusted measure rates was 0.62.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

HH AKI demonstrates robust score-level reliability, evaluated by Adams' SNR and ICC via the split-half sample approach. Specifically, Adams' SNRs ranged from 0.20 to 0.97 across the 20 test sites, with the mean and median equal to 0.84 and 0.91 respectively. Reliability may be low only for extremely small hospitals (e.g., one facility in our sample with only 150 eligible discharges). Analogously, the 100 estimated ICCs had a median of 0.99 and a mean ranging from 0.25 to 0.91. Risk adjustment, though introducing some variations to the average ICC estimates, did not impact the strength of score-level reliability. Overall, testing results clearly showed that HH AKI, as currently specified, can distinguish the true performance in AKI from one hospital from another.

[Response Ends]

2b. Validity

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

To empirically assess the data element validity, we compared data exported from the EHR (eData) to data manually abstracted from patients' medical charts (mData) for a subsample of measure initial population. We then calculated 1) the rate of missingness for all critical data elements needed for measure implementation, 2) percentage of match/agreement in each data element between the data sources (eData vs. mData), and 3) positive predictive value (PPV), sensitivity, negative predictive value (NPV), and specificity.

Manual abstraction is labor intensive; therefore, reducing burden while maximizing test result validity (e.g., level of power and significance) is important. We calculated the minimum required sample size (MRSS) using positive predictive value (PPV) as the primary endpoint and approximated MRSS using the conventional one-sample proportion formula, accounting for the intracluster correlation:[1] $n = \frac{z_{\frac{\alpha}{2}}^2 \cdot p \cdot (1-p)}{moe^2} \times VIF$ where αdenotes the type I error rate, moe denotes

the margin of error, p is PPV, and VIF is the variance inflation factor that accounts for the intracluster correlation. We simulated a series of $moes$, target ps, and the 95% confidence intervals associated with each p for different MRSS. Simulations indicated that with a moe of 6%, a target PPV of 0.9, a reasonable precision of PPV bounded by 0.84 and 0.96, and a conventionally accepted minimum number of observations that can render the sampling distribution of p to be normal, MRSS approximated 125.

We aimed for a random sample of 125 measure denominator cases for each of the three Meditech sites and 125 measure denominator cases across all 17 Cerner sites within one health system to enhance efficiency, given their similar clinical workflows and a shared central data warehouse. Manual chart review of patient medical records from any one site would thus be informative of records from other sites within the health system. We used random sampling without replacement. We additionally sampled 30 denominator-excluded encounters for each of the three Meditech sites and a separate sample of 30 cases across sites 2-18 to assess whether excluded cases per eData truly met the clinical intent for exclusion. Due to administrative and contractual delays, we ended up abstracting 157 cases for the 17 Cerner sites and 190 cases for each of the three Meditech sites, surpassing the MRSS.

To assess score-level validity, we applied known groups validity, one type of construct validity. Known groups validity focuses on the measure's ability to differentiate between groups of measured entities known to differ on their underlying latent construct. Prior research has suggested "known groups" that are identifiable using information that was available to testing:

- Hospital teaching/academic status
- Hospital bed size (<25, 25-99, 100-199, 200-499, and >499)
- Hospital urban/rural location

We hypothesize that the risk-adjusted AKI rate would be lower at teaching, large-sized, and urban hospitals than at non-teaching, small-sized, and rural hospitals, respectively. We dichotomized teaching versus non-teaching status because resident-to-bed ratios could not be linked at the facility level.

Footnotes:

1 What we mean by intracluster correlation here is a notion that hospitals with the same EHR system may have seen patients who are more alike. In this case, information revealed by patient A will not be entirely independent from that revealed by patient B. On the contrary, two sets of information share similarities and exhibit strong correlation. Without accounting for such intracluster correlation, we run the risk of underestimating sample size needed to yield a desired level of power and significance for the test statistics.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

Data Element Validity

Tables 4 to 11 show the percentage of match in terms of measure concepts, calculated using eData alone and separately calculated using mData alone. As above, we show results for the 17 Cerner sites combined and separately for each of the three Meditech sites.

Table 4. Percentage of Exact Match by Data Element; Denominator Specifications (Sites 1-17: Cerner)

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
Patient had an inpatient encounter with a discharge date between 1/1/20 and 12/31/20	127	127	100%
Patient aged ≥ 18 at the start of the encounter	127	127	100%
Inpatient length of stay ≥ 48hrs	127	127	100%
Patient had 1 or more serum creatinine lab result (not equal to zero) AFTER the first 48hrs of encounter start	127	127	100%
Patient had 2 or more serum creatinine lab results within the first 48hrs of encounter start	127	127	100%
Kidney Dialysis-Ordered or Performed **A physician order written (or kidney dialysis initiated) at greater than 48hrs after arrival to initiate kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis)**	4	4	100%
(Rolling) highest creatinine value 2.0 times or more of the (rolling) lowest creatinine value during the encounter	64	64	100%

Notes: hr = hour; CRRT = continuous renal replacement therapy.

Table 5. Percentage of Exact Match by Data Element; Denominator Specifications (Site 18: Meditech)

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
Patient had an inpatient encounter with a discharge date between 1/1/20 and 12/31/20	95	95	100%
Patient aged ≥ 18 at the start of the encounter	95	95	100%
Inpatient length of stay ≥ 48hrs	95	95	100%
Patient had 1 or more serum creatinine lab result (not equal to zero) AFTER the first 48hrs of encounter start	95	95	100%
Patient had 2 or more serum creatinine lab results within the first 48hrs of encounter start	95	95	100%
Kidney Dialysis-Ordered or Performed **A physician order written (or kidney dialysis initiated) at greater than 48hrs after arrival to initiate kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis)**	0	0	100%
(Rolling) highest creatinine value 2.0 times or more of the (rolling) lowest creatinine value during the encounter	6	6	100%

Notes: hr = hour; CRRT = continuous renal replacement therapy.

Table 6. Percentage of Exact Match by Data Element; Denominator Specifications (Site 19: Meditech)

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
Patient had an inpatient encounter with a discharge date between 1/1/20 and 12/31/20	99	99	100%
Patient aged ≥ 18 at the start of the encounter	99	99	100%
Inpatient length of stay ≥ 48hrs	99	99	100%
Patient had 1 or more serum creatinine lab result (not equal to zero) AFTER the first 48hrs of encounter start	99	99	100%
Patient had 2 or more serum creatinine lab results within the first 48hrs of encounter start	99	99	100%
Kidney Dialysis-Ordered or Performed **A physician order written (or kidney dialysis initiated) at greater than 48hrs after arrival to initiate kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis)**	10	10	100%
(Rolling) highest creatinine value 2.0 times or more of the (rolling) lowest creatinine value during the encounter	41	41	100%

Notes: hr = hour; CRRT = continuous renal replacement therapy.

Table 7. Percentage of Exact Match by Data Element; Denominator Specifications (Site 20: Meditech)

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
Patient had an inpatient encounter with a discharge date between 1/1/20 and 12/31/20	51	51	100%
Patient aged ≥ 18 at the start of the encounter	51	51	100%
Inpatient length of stay ≥ 48hrs	51	51	100%
Patient had 1 or more serum creatinine lab result (not equal to zero) AFTER the first 48hrs of encounter start	51	51	100%
Patient had 2 or more serum creatinine lab results within the first 48hrs of encounter start	51	51	100%
Kidney Dialysis-Ordered or Performed **A physician order written (or kidney dialysis initiated) at greater than 48hrs after arrival to initiate kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis)**	5	5	100%
(Rolling) highest creatinine value 2.0 times or more of the (rolling) lowest creatinine value during the encounter	18	19	95%

Notes: hr = hour; CRRT = continuous renal replacement therapy.

Table 8: Percentage of Exact Match by Data Element; Denominator Exclusions (Sites 1-17: Cerner)

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
AKI Present on Arrival Denominator Exclusion: Difference between the FIRST SCr value and any subsequent one collected within the first 48hrs of encounter start greater than or equal to 0.3 mg/dL	2	2	100%
ESRD CKD Stages 3, 4, or 5 Present on Arrival Denominator Exclusion: Patient had an eGFR ≤ 59 mL/min (using ONLY the non-Black value within the EHR) within the first 48-hours of encounter start	26	26	100%
ESRD Present on Arrival Denominator Exclusion: Patient had an ORDER for or had kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis) PERFORMED within the first 48 hours of encounter start	1	1	100%

Notes: hr = hour; SCr = serum creatinine; eGFR = estimated glomerular filtration rate; CRRT = continuous renal replacement therapy.

Table 9. Percentage of Exact Match by Data Element; Denominator Exclusions (Site 18: Meditech)

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
AKI Present on Arrival Denominator Exclusion: Difference between the FIRST SCr value and any subsequent one collected within the first 48hrs of encounter start greater than or equal to 0.3 mg/dL	8	8	100%
ESRD CKD Stages 3, 4, or 5 Present on Arrival Denominator Exclusion: Patient had an eGFR ≤ 59 mL/min (using ONLY the non-Black value within the EHR) within the first 48 hours of encounter start	89	89	100%
ESRD Present on Arrival Denominator Exclusion: Patient had an ORDER for or had kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis) PERFORMED within the first 48 hours of encounter start	1	1	100%

Notes: hr = hour; SCr = serum creatinine; eGFR = estimated glomerular filtration rate; CRRT = continuous renal replacement therapy.

Table 10. Percentage of Exact Match by Data Element; Denominator Exclusions (Site 19: Meditech)

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
AKI Present on Arrival Denominator Exclusion: Difference between the FIRST SCr value and any subsequent one collected within the first 48hrs of encounter start greater than or equal to 0.3 mg/dL	12	12	100%

Data Element	No. of cases per EHR	No. of cases per Abstraction	Percent Match (%)
ESRD CKD Stages 3, 4, or 5 Present on Arrival Denominator Exclusion: Patient had an eGFR ≤ 59 mL/min (using ONLY the non-Black value within the EHR) within the first 48 hours of encounter start	88	88	100%
ESRD Present on Arrival Denominator Exclusion: Patient had an ORDER for or had kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis) PERFORMED within the first 48 hours of encounter start	5	5	100%

Notes: hr = hour; SCr = serum creatinine; eGFR = estimated glomerular filtration rate; CRRT = continuous renal replacement therapy.

Table 11. Percentage of Exact Match by Data Element; Denominator Exclusions (Site 20: Meditech)

Data Element	No. of cases per the EHR	No. of cases per Abstraction	Percent Match (%)
AKI Present on Arrival Denominator Exclusion: Difference between the FIRST SCr value and any subsequent one collected within the first 48hrs of encounter start greater than or equal to 0.3 mg/dL	47	47	100%
ESRD CKD Stages 3, 4, or 5 Present on Arrival Denominator Exclusion: Patient had an eGFR ≤ 59 mL/min (using ONLY the non-Black value within the EHR) within the first 48 hours of encounter start	125	125	100%
ESRD Present on Arrival Denominator Exclusion: Patient had an ORDER for or had kidney dialysis (CRRT, hemodialysis, or peritoneal dialysis) PERFORMED within the first 48 hours of encounter start	21	21	100%

Notes: hr = hour; SCr = serum creatinine; eGFR = estimated glomerular filtration rate; CRRT = continuous renal replacement therapy.

Tables 12 to **15** present findings for PPV, sensitivity, NPV, and specificity. For the measure numerator, PPV denotes the probability that an EHR-reported HA-AKI is a valid HA-AKI based on the clinical review of patients' medical records. For measure denominator exclusions, PPV denotes the probability that cases excluded from the measure per the EHR truly met the clinical rationale for exclusion.

Table 12. PPV, Sensitivity, NPV, Specificity; Sites 1-17: Cerner

Measure IPP/Denom Exclsn/Denom-only/Numerator	Per Abstraction	Per the EHR	PPV	Sensitivity	NPV	Specificity
Initial population	157	157	100%	100%	100%	100%
Denominator exclusion	38	30	100%	79%	94%	100%
Numerator negative	61	63	97%	100%	100%	97.9%

Measure IPP/Denom Exclsn/Denom-only/Numerator	Per Abstraction	Per the EHR	PPV	Sensitivity	NPV	Specificity
Numerator	58	64	90.6%	100%	100%	93.9%

Notes: IPP = initial population; Exclsn = exclusion; Numerator negative = encounters in denominator but not in numerator; PPV = positive predicative value; NPV = negative predicative value.

Table 13. PPV, Sensitivity, NPV, Specificity; Site 18: Meditech

Measure IPP/Denom Exclsn/Denom-only/Numerator	Per Abstraction	Per the EHR	PPV	Sensitivity	NPV	Specificity
Initial population	190	190	100%	100%	100%	100%
Denominator exclusion	94	95	98.9%	100%	100%	99.0%
Numerator negative	90	89	100%	99%	99.0%	100%
Numerator	6	6	100%	100%	100%	100%

Notes: IPP = initial population; Exclsn = exclusion; Numerator negative = encounters in denominator but not in numerator; PPV = positive predicative value; NPV = negative predicative value.

Table 14. PPV, Sensitivity, NPV, Specificity; Site 19: Meditech

Measure IPP/Denom Exclsn/Denom-only/Numerator	Per Abstraction	Per the EHR	PPV	Sensitivity	NPV	Specificity
Initial population	190	190	100%	100%	100%	100%
Denominator exclusion	91	91	100%	100%	100%	100%
Numerator negative	49	49	100%	100%	100%	100%
Numerator	50	50	100%	100%	100%	100%

Notes: IPP = initial population; Exclsn = exclusion; Numerator negative = encounters in denominator but not in numerator; PPV = positive predicative value; NPV = negative predicative value.

Table 15. PPV, Sensitivity, NPV, Specificity; Site 20: Meditech

Measure IPP/Denom Exclsn/Denom-only/Numerator	Per Abstraction	Per the EHR	PPV	Sensitivity	NPV	Specificity
Initial population	190	190	100%	100%	99.4%	100%
Denominator exclusion	139	139	100%	100%	100%	100%
Numerator negative	27	28	96%	100%	100%	99.4%
Numerator	24	23	100%	96%	99.4%	100%

Notes: IPP = initial population; Exclsn = exclusion; Numerator negative = encounters in denominator but not in numerator; PPV = positive predicative value; NPV = negative predicative value.

Table 16 shows the summary statistics for known-groups validity. On average, teaching hospitals performed better (27%) than non-teaching hospitals (1.87% versus 1.37%), and large and urban hospitals performed better than small and rural hospitals (e.g., 2.6% for hospitals with <100 beds versus 1.1% for hospitals with >499 beds). Due to small sample size, these testing results are not conclusive.

Table 16. Known Groups Validity for the AKI eCQM

Known Groups Category	N	Mean	Std.Dev
Academic status	*	*	*
Non-teaching	15	0.0187	0.0100
Teaching	5	0.0137	0.0033
Urban vs. Rural	*	*	*
Rural	15	0.0208	0.0133
Urban	5	0.0163	0.0074
Bed size	*	*	*
<25	1	0.0257	N/A
25-99	3	0.0259	0.0156
100-199	8	0.0182	0.0084
200-499	7	0.0127	0.0035
>499	1	0.0113	N/A

Note: Hospital-level measure rate reflects risk adjustment. Std.Dev = standard deviation. N/A = not applicable. *Cells intentionally left empty.

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]

Testing results show that the percentage of match in measure concepts between data sources was nearly perfect across sites. Results also show robust PPVs for every measure component across sites. For instance, PPV ranges from 90.6% to 100% when comparing measure numerator between data sources. PPVs for the other measure components are generally high, with a low of 96% and a mode equal to 100%. Overall, testing results point to strong data element level validity and indicate that the measure's ability to correctly classify AKI among patients who experienced AKI is high.

Testing results on the measure's known-groups validity (score-level validity) are consistent with our expectations. For example, teaching hospitals' risk-adjusted AKI rates were 27% lower on average than non-teaching hospitals' risk-adjusted AKI rates, and large hospitals performed better than small hospitals. Descriptive findings suggest that large hospitals have better resources and staff skill mix than small hospitals, but the findings are not conclusive due to the relatively small number of facilities.

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

Using the full denominator data, we first calculated the hospital-level measure rate and its 95% confidence interval for each of the 20 test sites. We then calculated the system-wide, weighted average measure rate across 20 test sites. Lastly, we compared each test hospital's performance against the system-wide average and assessed whether its performance differs significantly from the weighted mean.

Alternatively, we ran two linear regressions, relating the incidence of AKI to a set of hospital-specific fixed effects and estimating a generalized T-test. In the first regression, we did not adjust for patient characteristics, while in the second regression we did.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Exhibit 3 shows the distribution of observed hospital performance rate and its 95% confidence interval. It also shows the system-wide, weighted average calculated across the 20 test sites.

Exhibit 3. Distribution of Measure Performance Rate Across Test Sites; Data in CY2020

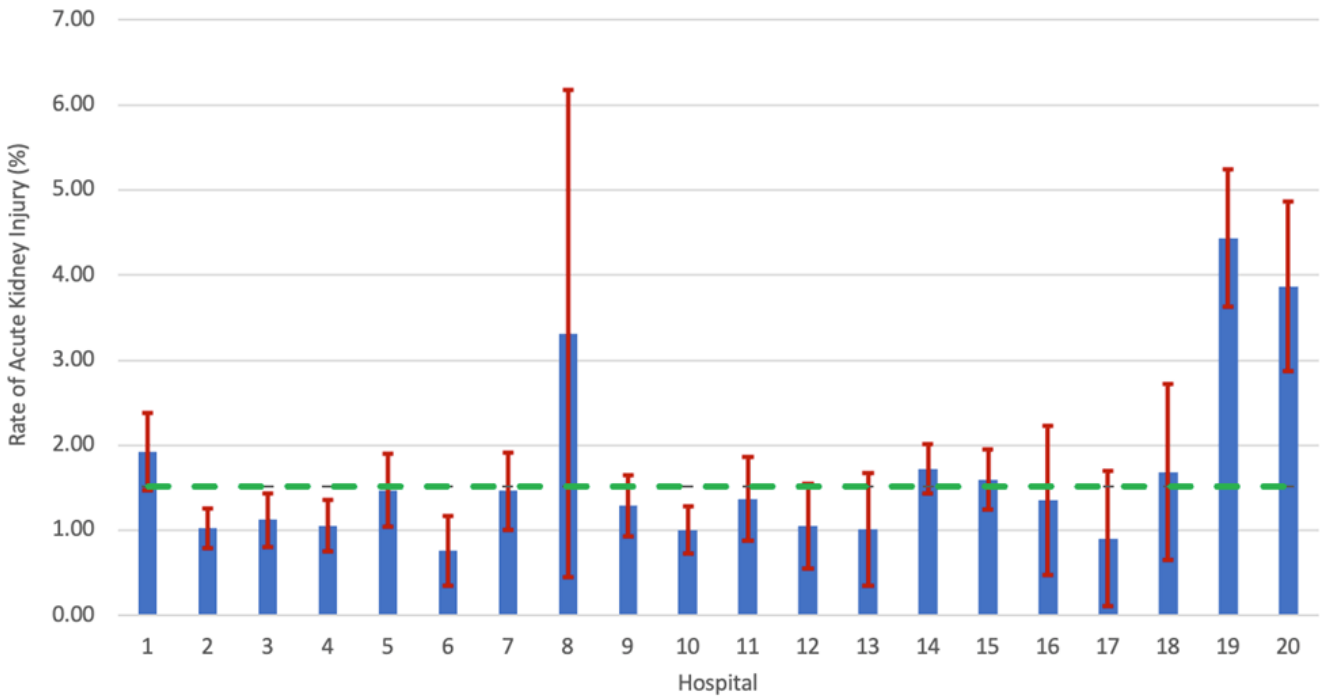


Table 17 below shows regression coefficients and their cluster-robust 95% confidence intervals for the alternative analysis described above.

Table 17. Regression Coefficients and 95% Confidence Intervals

*	**Without controlling for patient-mix**	**Controlling for patient-mix**
*	Coef./95% CI	Coef./95% CI
Test site 1 (ref grp)	*	*
*	*	*
Test site 2	-0.009***	-0.007***
*	[-0.009,-0.009]	[-0.008,-0.006]
Test site 3	-0.008***	-0.008***
*	[-0.008,-0.008]	[-0.009,-0.007]
Test site 4	-0.009***	-0.005***
*	[-0.009,-0.009]	[-0.006,-0.005]
Test site 5	-0.004***	-0.004***
*	[-0.004,-0.004]	[-0.004,-0.003]
Test site 6	-0.012***	-0.008***
*	[-0.012,-0.012]	[-0.009,-0.007]
Test site 7	-0.005***	-0.002***
*	[-0.005,-0.005]	[-0.002,-0.001]
Test site 8	0.014***	0.016***
*	[0.014,0.014]	[0.015,0.017]
Test site 9	-0.006***	-0.010***
*	[-0.006,-0.006]	[-0.011,-0.009]
Test site 10	-0.009***	-0.011***
*	[-0.009,-0.009]	[-0.012,-0.011]
Test site 11	-0.006***	-0.004***
*	[-0.006,-0.006]	[-0.004,-0.004]
Test site 12	-0.009***	-0.004***
*	[-0.009,-0.009]	[-0.005,-0.002]
Test site 13	-0.009***	-0.006***
*	[-0.009,-0.009]	[-0.006,-0.005]
Test site 14	-0.002***	-0.010***
*	[-0.002,-0.002]	[-0.012,-0.009]
Test site 15	-0.003***	-0.006***

*	Without controlling for patient-mix	Controlling for patient-mix
*	[-0.003,-0.003]	[-0.007,-0.006]
Test site 16	-0.006***	0.002**
*	[-0.006,-0.006]	[0.000,0.003]
Test site 17	-0.010***	-0.009***
*	[-0.010,-0.010]	[-0.010,-0.009]
Test site 18	-0.002***	0.000
*	[-0.002,-0.002]	[-0.000,0.001]
Test site 19	0.025***	0.020***
*	[0.025,0.025]	[0.018,0.022]
Test site 20	0.019***	0.014***
*	[0.019,0.019]	[0.013,0.016]
Observations	58,936	58,936

Notes: Regressions run using all data points from the measure denominator population. Ref grp = reference group. Standard errors clustered at the level of hospital. *** p < 0.01. *Cells intentionally left empty.

[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

Measure performance rates ranged from 0.76% to 4.43%, indicating ample room for quality improvement in the inpatient setting. Several hospitals' performance rates are consistently below the system-wide average while a few others are above that mean. Regression results, complementing the bar graph, confirm that the measure can identify clinically meaningful differences in AKI across hospitals.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

We assessed the magnitude of data missingness during the process of parallel-form comparison, as described in detail in section 2b.02. Samples were drawn for abstraction based on random sampling without replacement.

During the abstraction, we compared data exported from the EHR (eData) to data manually abstracted from patients' medical charts (mData) for every patient included in the abstraction sample. Given that information in patients' medical

charts is typically deemed the "gold standard," this process helped us to identify the extent and distribution of missing data and assess in what direction measure performance calculated using the EHR data could be biased if data missing is systematic. We tabulated the frequency and percentage of data missingness for each key data element.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

Tables 18-21 below show the frequency of missing data for the critical data elements for the 17 Cerner sites combined and for each of the three Meditech sites, respectively. Exclusions based on ICD-10-CM or ICD-10-PCS diagnosis codes were also validated, and no discrepancies were found.

Table 18. Missing Data for Critical Data Elements (Cerner Sites 1-17)

Data Element	Missing Count	Total Count Per EHR	Total Count Per Abstraction	Missing Percent (%)
Patient encounter (ED, OBS, or IP) Start DateTime	0	157	157	0%
Patient encounter (IP) End DateTime	0	157	157	0%
Patient birth date	0	157	157	0%
Patient serum creatinine value	0	2,005	2,005	0%
Patient serum creatinine DateTime	0	2,005	2,005	0%
Patient eGFR value	0	467	467	0%
Patient eGFR DateTime	0	467	467	0%
Kidney dialysis ordered or performed DateTime	4	59	63	6%

Table 19. Missing Data for Critical Data Elements (Meditech Site 18)

Data Element	Missing Count	Total Count Per EHR	Total Count Per Abstraction	Missing Percent (%)
Patient encounter (ED, OBS, or IP) Start DateTime	0	190	190	0%
Patient encounter (IP) End DateTime	0	190	190	0%
Patient birth date	0	190	190	0%

Data Element	Missing Count	Total Count Per EHR	Total Count Per Abstraction	Missing Percent (%)
Patient serum creatinine value	3	1,170	1,173	0.3%
Patient serum creatinine DateTime	3	1,170	1,173	0.3%
Patient eGFR value	44	1,120	1,164	3.8%
Patient eGFR DateTime	44	1,120	1,164	3.8%
Kidney dialysis ordered or performed DateTime	0	24	24	0%

Table 20. Missing Data for Critical Data Elements (Meditech Site 19)

Data Element	Missing Count	Total Count Per EHR	Total Count Per Abstraction	Missing Percent (%)
Patient encounter (ED, OBS, or IP) Start DateTime	0	190	190	0%
Patient encounter (IP) End DateTime	0	190	190	0%
Patient birth date	0	190	190	0%
Patient serum creatinine value	0	1,894	1,894	0%
Patient serum creatinine DateTime	0	1,894	1,894	0%
Patient eGFR value	0	1,894	1,894	0%
Patient eGFR DateTime	0	1,894	1,894	0%
Kidney dialysis ordered or performed DateTime	0	116	116	0%

Table 21. Missing Data for Critical Data Elements (Meditech Site 20)

Data Element	Missing Count	Total Count Per EHR	Total Count Per Abstraction	Missing Percent (%)
Patient encounter (ED, OBS, or IP) Start DateTime	0	190	190	0%
Patient encounter (IP) End DateTime	0	190	190	0%
Patient birth date	0	190	190	0%
Patient serum creatinine value	0	1,768	1,768	0%
Patient serum creatinine DateTime	0	1,768	1,768	0%

Data Element	Missing Count	Total Count Per EHR	Total Count Per Abstraction	Missing Percent (%)
Patient eGFR value	0	1,767	1,767	0%
Patient eGFR DateTime	0	1,767	1,767	0%
Kidney dialysis ordered or performed DateTime	0	218	218	0%

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

Testing results clearly show that all critical data elements are consistently stored in the EHR and can be accurately exported for calculation. For example, the rate of data missingness was zero across test sites except for a modest number of eGFR values that weren't captured in the EHR at one test site. This problem was resolved by applying the revised (race-neutral) CKD-EPI formula to serum creatinine values (which were almost never missing) to calculate eGFR consistently across all observations at all sites. Therefore, any bias due to missing eGFR values was virtually eliminated.

Missingness of "kidney dialysis ordered or performed" was attributed to the use of third-party contracted services to perform dialysis at a few sites, leading to documentation of dialysis only in unstructured form (e.g., scanned paper). This problem can be resolved by facilities' ensuring that the performance of dialysis is documented in structured fields, a small workflow improvement.

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

We used two methods to test the empirical impact of measure denominator exclusions. First, using the full denominator data, we removed measure exclusion criterion one at a time from the logic and calculated the marginal and relative (%) impact on the prevalence of the numerator and denominator, as well as the observed measure rate. Second, through parallel-form comparison, we evaluated whether patients excluded from the denominator per the EHR truly met the clinical intent for exclusion.

[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

2022 NQF ITS Attachment: Sheet 4 provides the full test results, stratified by test site, for the first method. **Table 22** below shows the results by pooling together all 20 test sites.

Table 22. Impact of Denominator Exclusion Criteria on Denominator Count, Numerator Count, and Measure Rate; All Test Sites Combined

Acute Kidney Injury	Denominator Count	Denominator %change	Numerator Count	Numerator %change	Rate per 100 Dropping Exclsn	Rate per 100 %change
Current specification	58,936	*	896	*	1.52	*
Remove: At least 2 SCrs within 48hrs of Encounter Start	62,815	6.58	939	4.80	1.49	-1.67
Remove: Increase of 0.3 or greater in SCr relative to Index SCr within 48hrs of Encounter Start	61,269	3.96	1,288	43.75	2.10	38.28
Remove: Index eGFR^ < 60 mL/min	81,230	37.83	1,576	75.89	1.94	27.62
Remove: Dialysis Ordered or Performed within 48hrs of Encounter Start	58,949	0.02	898	0.22	1.52	0.20
Remove: ICD-10-CM/PCS based Exclusions**	60,045	1.88	921	2.79	1.53	0.89

Note: *Cells intentionally left empty. ^eGFR is based on the new 2021 CKD-EPI creatinine formula in Inker et al. (2021). **ICD-10-CM/PCS based exclusions encompass Thrombotic thrombocytopenic purpura, Rapidly Progressive Nephritic Syndrome, Traumatic Avulsion, Hemolytic Uremic Syndrome (HUS), Large Body Surface Area (BSA) Burns, Obstetrics, Obstetrics VTE, ECMO, IABP, REBOA, and Nephrectomy. SCr = serum creatinine. Hrs = hours. eGFR = estimated glomerular filtration rate. Exclsn = exclusion.

Please refer to section 2b.03 for the test results for the second method.

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]

Across the five exclusion categories (all ICD-10-CM [PCS] based exclusions are grouped under one category) and the 20 test sites, the average relative impact on the denominator, numerator, and measure rate as a result ranged from less than 1% to approximately 40%, less than 1% to more than 75%, and negative 2% to nearly 40%, respectively. We now turn to each exclusion criterion to describe how it successfully removes records that would otherwise be falsely labeled as HH AKI.

1. **At Least Two Serum Creatinine Within 48 Hours of Encounter Start**. This is a necessary criterion to determine whether patients have AKI or moderate-to-severe CKD on arrival. **Table 22** indicates that removing this criterion increased the total denominator across sites by 6.6% and total numerator by 4.8%. The resulted measure rate across sites decreased slightly from 1.52% to 1.49%, which amounts to a 1.7% relative reduction.
2. **Increase of 0.3 mg/dL or Greater in Serum Creatinine Value Relative to the Index Serum Creatinine within 48 Hours of Encounter Start**. This condition uses objective test results to exclude patients who have AKI at the start of care, due to underlying conditions such as dehydration, hemorrhage, traumatic injury, or sepsis. We define index serum creatinine as the lowest serum creatinine value within 24 hours of encounter start (inclusive of emergency department visit and observation stay) or the very first value within 48 hours of encounter start if there is no serum creatinine during the first 24 hours. **Table 22** shows that relaxing this criterion increased the total denominator by roughly 4% but the total numerator by nearly 44%. Clinical knowledge and empirical findings confirmed that this criterion is integral to reducing the measure's false positive rate.
3. **Index eGFR < 60 mL/min**. This criterion complements the above two and excludes patients who had moderate-to-severe CKD at the start of care. **Table 22** shows that removing this criterion from measure logic made a substantial impact on both denominator (38% increase) and numerator (76% increase). Empirical findings support the clinical understanding that these patients may falsely appear to have AKI due to progression of their underlying CKD or nonpreventable changes in volume status.
4. **Dialysis Performed Within 48 Hours of Encounter Start.** This exclusion criterion uses the initiation of hemodialysis or peritoneal dialysis to remove patients with end stage renal disease, AKI, or severe acute metabolic derangements at the start of the encounter from the measure denominator. **Table 22** shows that removing this criterion had minimal impact on the denominator, numerator, and observed rate. Although criteria 2 and 3 capture nearly all patients who have AKI at presentation, this criterion is useful at the margin because dialysis may occasionally be necessary for patients with index eGFR values of 60 or greater.
5. **ICD-10-CM (PCS) Based Complications**. This exclusion category utilizes ICD-10-CM diagnosis with POA indicators and ICD-10-PCS codes to identify patients with extremely high-risk conditions or procedures and remove them from the measure denominator. **Table 22** suggests that removing these code-based exclusions increased the total denominator and numerator by roughly 2% and 3%, respectively. These exclusions are clinically well justified. For example, patients who undergo nephrectomy have a very high likelihood of developing AKI, yet nephrectomy is often necessary to treat renal tumors.[1,2]

Note that HH AKI does not consider obstetric patients in hospital performance evaluation because there is an existing measure in CMS's quality programs that focuses exclusively on obstetric patients. As a result, including this subpopulation becomes unnecessary. Overall, all exclusions are necessary to reduce the measure's false positive rate and to exclude patients for whom clinical experts agreed that AKI is essentially nonpreventable. None of the denominator exclusions imposes a burden on providers by increasing the complexity of data collection or analysis because all data exist in the EHR and are already collected during a part of routine care.

References:

1. Lee, Y., Ryu, J., Kang, M. W., Seo, K. H., Kim, J., Suh, J., ... & Han, S. S. (2021). Machine learning-based prediction of acute kidney injury after nephrectomy in patients with renal cell carcinoma. *Scientific reports*, *11*(1), 1-8.
2. Kaushik, D., Kim, S. P., Childs, M. A., Lohse, C. M., Costello, B. A., Cheville, J. C., ... & Thompson, R. H. (2013). Overall survival and development of stage IV chronic kidney disease in patients undergoing partial and radical nephrectomy for benign renal tumors. *European urology*, *64*(4), 600-606.

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

Statistical risk model with risk factors (specify number of risk factors)

[Statistical risk model with risk factors (specify number of risk factors) Please Explain]

A total of 28 risk coefficients from 13 distinct risk factors was used in the risk model.

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

Guided by a streamlined conceptual model, we developed the baseline risk adjustment model for HH AKI using a five-step process.

1. Randomly partitioned the full denominator data into a training set, a validation set, and a holdout (model performance evaluation) set, with data splitting stratified by hospital and outcome.
2. Ran locally weighted regression, relating measure outcome to each candidate risk factor, and cross-tabulated each factor against measure outcome to examine risk factors' functional forms.
3. Fit the initial model using the least absolute shrinkage and selection operator (LASSO) on the training set and subsequently examined model fit on the validation set. Validation set helped to assess model fit on the training set, while facilitating parameter tuning (e.g., the lambda regularization parameter in the cross-validation [CV]-based LASSO). This step is iterative, and we considered three common approaches in parameter tuning: minimum of CV function, "one-standard-error" (or one SE) rule promoted by Hastie, Tibshirani, and Wainwright (2015),[1] and minimum of Bayesian Information Criterion (BIC) function suggested by Zhang, Li, and Tsai (2010).[2]
4. Compared selected features (or risk factors) across the three methods, assessed each risk factor's directional impact on the outcome in the training set via a standard logistic regression, and brought in clinical expertise to finalize features suited for the baseline risk model.
5. Estimated a final, unbiased assessment of model performance on the holdout set. Since data points in the holdout set were not used for model training or parameter tuning, the model design (i.e., feature selection and parameter tuning) in this stage remains unchanged.

Baseline risk adjustment model is shown below in **Table 24**, with risk factor names, denominator prevalences, coefficient (logit) estimates, and their cluster-robust 95% confidence intervals. Risk factors used in the baseline model are:

- Patient sex-by-age category interactions. Patient sex is either male or female. Patient age is grouped into four bins: 18-34, 35-54, 55-74, and 75 or above. Female-by-Age(18-34) is the reference group.
- Vital signs at the encounter start are classified based on the APACHE II system and further collapsed.[3]
    - Temperature has two categories: abnormally low temperature (< 36°C) and normal or high temperature (≥ 36°C). Normal or high temperature is the reference group.
    - Heart rate has two categories: low or normal heart rate (≤ 109 beats per minute) and abnormally high heart rate (> 109 beats per minute). Low or normal heart rate is the reference group.
    - Respiratory rate has three categories: abnormally low respiratory rate (< 12 breaths per minute), normal respiratory rate (≥ 12 but < 24 breaths per minute), and abnormally high respiratory rate (>24 breaths per minute). Normal respiratory rate is the reference group.
    - Systolic blood pressure (SBP) is grouped into three bins: < 110 (mm Hg), ≥ 110 but ≤ 160 (mm Hg), and > 160 (mm Hg). SBP ≥ 110 but ≤ 160 is the reference group.
- Index eGFR. We derived the index eGFR based on the index serum creatinine lab value, patient sex, and patient age using formula specified in Inker et al. (2021).[4] We winsorized the index eGFR at 120 to minimize the impact of artificially high eGFR among the elderly (due to their greatly reduced muscle bulk).
- AHRQ (Elixhauser) comorbidities[5]
    - Cancer (leukemia, lymphoma, or metastatic cancer)
    - Diabetes
    - Heart failure
    - Hypertension
    - Obesity
- Encounter length of stay:
    - 0 – 7 days (reference group)

- 7 – 14 days
- 15 – 21 days
- 21 – 30 days (Note that measure calculation is capped at 30 days since the encounter start)

All risk factors came from the EHR, including comorbidities implemented using external software (e.g., AHRQ's Elixhauser Comorbidity Software).

Table 23 below shows each risk factor's associated code system and codes.

Table 23. Associated code system and codes used in the risk adjustment model for eCQM Hospital Harm - Acute Kidney Injury

Risk Adjustment Variables	Associated Code System and Codes
Sex and age	HL7V3.0_2019-12
Vital Signs	*
Temperature	LOINC: 76011-6; 8310-5; 8328-7; 8329-5; 8331-1,8332-9; 8333-7
Heart Rate	LOINC: 8867-4
Respiratory Rate	LOINC: 9279-1
Systolic Blood Pressure	LOINC: 8480-6
Index estimated glomerular filtration rate (eGFR)^	LOINC: 21232-4; 2160-0; 38483-4
Elixhauser Comorbidities**	*
Cancer (leukemia, lymphoma, or metastatic cancer)	ICD-10-CM: See 2022 NQF ITS Attachment: Sheet 3
Diabetes†	ICD-10-CM: See 2022 NQF ITS Attachment: Sheet 3
Heart failure	ICD-10-CM: See 2022 NQF ITS Attachment: Sheet 3
Hypertension††	ICD-10-CM: See 2022 NQF ITS Attachment: Sheet 3
Obesity	ICD-10-CM: See 2022 NQF ITS Attachment: Sheet 3
Encounter Length of stay (LOS)	SNOMEDCT: 4525004
Encounter Length of stay (LOS)	SNOMEDCT: 448951000124107
Encounter Length of stay (LOS)	SNOMEDCT: 183452005; 32485007; 8715000

Notes:

* Cells left intentionally empty.

^ The measure derives the index eGFR using formula in Inker et al. (2021)[4] based on the index serum creatinine value and patient sex and age. LOINC codes listed on the right show the code system and the associated codes for serum creatinine. **Elixhauser Comorbidity Software Refined for ICD-10-CM, v2022.1 can be found at https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp.† Diabetes include diabetes with chronic complications and diabetes without chronic complications.

†† Hypertension includes hypertension, complicated and hypertension, uncomplicated

Upon selecting risk factors, we fit the baseline risk model using the standard approach that accounts for the clustering of patients within hospitals but accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes. The model specification uses a hierarchical logistic regression, with the assumption that the conditional distribution of AKI given hospital-level random effects is Bernoulli distributed, and the event probability is determined by a logistic cumulative distribution function. $Pr(y_{ij} = 1|x_{ij}, u_j) = H(x_{ij}\beta + z_{ij}u_j)$ where $j = 1, 2, \ldots, M$ clusters (hospitals), with hospital j consisting of $i = 1, 2, \ldots, n_j$ observations (qualified

inpatient encounters). The outcome variables (responses) are binary-valued y_{ij}, and we define $y_{ij} = 1$ if AKI occurred for inpatient encounter i in hospital j and 0 otherwise. The $1 \times p$ vector X_{ij} subsumes the encounter-level risk factors, with coefficients β. The $1 \times q$ vector z^{ij} represents the random intercepts. Of note, we do not include any random-level factors (e.g., bed size, FTE residents, FTE RNs) since they are under hospitals' control, z^{ij} is thus a scalar of 1. Without loss of generality, we retain z^{ij} in the specification. The random effects u_j are M realizations from a multivariate normal distribution with mean 0 and $q \times q$ variance matrix$\sum$, which can be viewed as the between-hospital variance in the quality of care. Random effects will not be directly estimated as model parameters in the absence of hospital-level factors. Instead, they are summarized according to the variance components $\sum$.

By construction, $H(\cdot)$ is the logistic cumulative distribution function that maps risk factors to the probability of AKI $(y_{ij} = 1)$, with $H(v) = \frac{exp(v)}{1+exp(v)}$.

Table 24. HH AKI Baseline Risk-Adjustment Model: Risk Factors with Prevalencies, Parameter Estimates, and 95% Confidence Intervals.

Risk Factor	Denom Ct.	Logit Coef.	Robust S.E Clustered at Hospital	95% Confidence Interval
Intercept	N/A	-5.534	0.429	(-6.376, -4.693)
Sex-by-Age groups	*	*	*	*
Female×Age (18-34) – Ref.	2,482	–	–	–
Female×Age (35-54)	6,533	0.079	0.309	(-0.528, 0.686)
Female×Age (55-74)	11,479	-0.221	0.285	(-0.778, 0.337)
Female×Age (75 and older)	7,380	-0.238	0.284	(-0.795, 0.319)
Male×Age (18-34)	3,023	0.049	0.306	(-0.550, 0.649)
Male×Age (35-54)	7,963	-0.047	0.271	(-0.579, 0.485)
Male×Age (55-74)	13,887	-0.179	0.312	(-0.790, 0.433)
Male×Age (75 and older)	6,189	-0.178	0.3	(-0.766, 0.410)
Vital Signs	*	*	*	*
Abnormally low Temp	1,885	0.246	0.14	(-0.029, 0.520)
Normal or high Temp – Ref.	57,051	–	–	–
Low or normal HR – Ref.	45,155	–	–	–
Abnormally high HR	13,781	0.263	0.108	(0.050, 0.475)
Abnormally low Resp rate	538	0.287	0.182	(-0.069, 0.643)
Normal Resp rate – Ref.	52,435	–	–	–
Abnormally high Resp rate	5,963	0.214	0.081	(0.055, 0.372)
SBP < 110	8,110	0.053	0.126	(-0.194, 0.299)
SBP ≥ 110 & ≤ 160 – Ref.	40,096	–	–	–
SBP > 160	10,730	0.259	0.102	(0.060, 0.459)
Index eGFR	N/A	-0.007	0.003	(-0.013, -0.001)
Elixhauser Comorbidities	*	*	*	*

Risk Factor	Denom Ct.	Logit Coef.	Robust S.E Clustered at Hospital	95% Confidence Interval
Cancer LLM	4,076	0.582	0.103	(0.380, 0.784)
Diabetes	18,556	0.164	0.087	(-0.006, 0.334)
Heart failure	9,945	0.542	0.085	(0.375, 0.708)
Hypertension	36,890	0.279	0.084	(0.114, 0.444)
Obesity	13,416	0.264	0.079	(0.108, 0.419)
Encounter LOS	*	*	*	*
>0d & ≤ 7d – Ref.	43,492	–	–	–
>7d & ≤14d	11,959	2.148	0.104	(1.944, 2.353)
>14d & ≤21d	2,196	3.348	0.094	(3.164, 3.532)
>21d & ≤30d	1,289	4.136	0.153	(3.836, 4.436)
Variance component	N/A	0.147	0.052	(0.074, 0.293)

Notes: Coef. = coefficient; Ct. = count; Denom = denominator; eGFR = estimated glomerular filtration rate; HR = heart rate; LLM = leukemia, lymphoma, or metastatic cancer; LOS = length of stay; Resp = respiratory; SBP = systolic blood pressure; S.E = standard errors; Temp = temperature; Ref. = reference group.

*Cells intentionally left empty.

References:

1. Hastie, T., Tibshirani, R., & Wainwright, M. (2019). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC.
2. Zhang, Y., Li, R., & Tsai, C. L. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*, *105*(489), 312-323.
3. Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Critical Care Medicine*, *13*(10), 818-829.
4. Inker, L. A., Eneanya, N. D., Coresh, J., Tighiouart, H., Wang, D., Sang, Y., ... & Levey, A. S. (2021). New creatinine- and cystatin C–based equations to estimate GFR without race. New England Journal of Medicine, 385(19), 1737-1749.
5. AHRQ Elixhauser Comorbidity Software Refined for ICD-10-CM (version 2022.1), publicly available at https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp

[Response Ends]

Attachment: 3713e_3713e_2022_NQF_ITS_Attachment_To NQF_(6)-508.xlsx

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Published literature

Internal data analysis

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

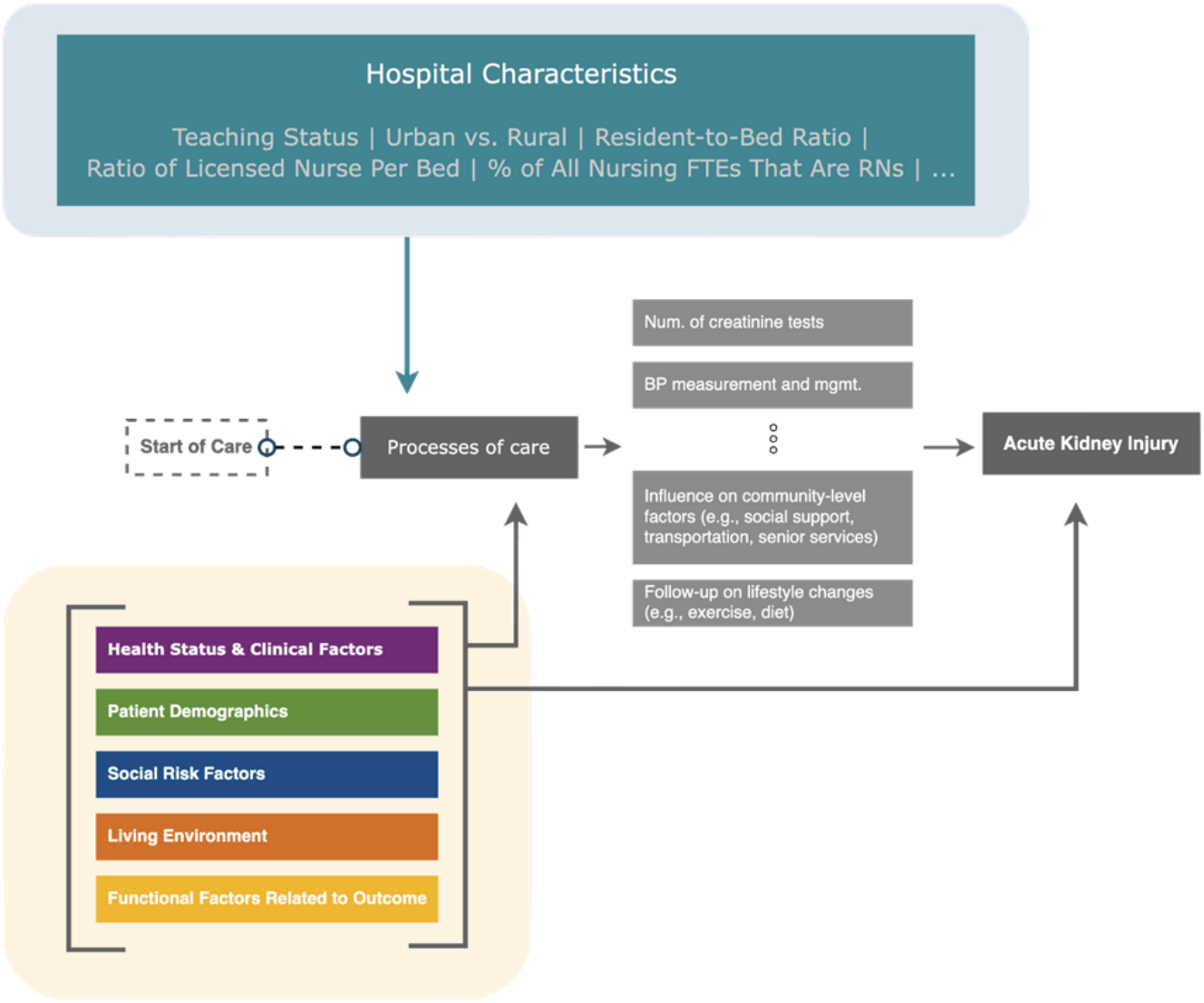
Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

AKI is a major risk factor for the development of CKD. According to the United States Renal Data System, acute tubular necrosis without recovery is the primary diagnosis for 2-3% of incident end-stage renal disease cases (ESRD) annually.[1] Other studies have demonstrated significantly increased long-term risk of CKD and ESRD following AKI, even after initial recovery of kidney function.[2] Annually, AKI affects up to 10% of hospitalized patients, a rate that is comparable to severe sepsis and acute lung injury.[3] While some instances of AKI are non-preventable, a proportion of them can be treated and avoided using best practices in care. For example, the KDIGO guidelines suggest that careful management of hemodynamic status, fluids, and vasoactive medications prevent the occurrence of AKI during hospitalization.[4]

Exhibit 4 gives a simplified conceptual framework that guided the development of our baseline risk model. Hospital characteristics, such as teaching/academic status, nurse skill mix, and urban/rural location, may affect the processes of care implemented and quality of care provided. Patient characteristics, such as sex, age, baseline kidney function, and comorbid conditions, are independently associated with the risk of AKI. Mediators, sitting between patient characteristics and the development of AKI, represent the pathways by which improved care can lead to better outcomes.

Exhibit 4. Simplified Conceptual Model That Guided the Risk Adjustment Model Development



Note: The simplified conceptual model contains three components that are correlated with the measured outcome of interest (AKI). The top part shows hospital characteristics, such as academic/teaching status and the bottom part shows patient characteristics (age, sex, clinical symptoms, etc.) that are already present at the start of care. Mediating factors, such as hospital's influence on patients' living environment, are located between the processes of care and measured outcome of interest. FTE stands for full-time equivalent. RN stands for registered nurse. BP stands for blood pressure. Mgmt. stands for management.

Patient Characteristics: Patient attributes (demographics, comorbid conditions, clinical signs and symptoms, social and functional risks, and others) present at the start of care directly influence the measured outcome and hospitals have little control. It is, however, important to recognize that patient characteristics may exert indirect effects that are mediated by quality-of-care factors:

1. Patients of lower income/education or unstable housing may not have equitable access to high-quality facilities when these facilities choose to avoid high-poverty areas. As a result, economically disadvantaged patients may be more likely to visit lower-quality hospitals, which can contribute to their increased risk of AKI during hospitalization.[5,6]
2. Patients may not receive equivalent care within the hospital when care rationing exists. For example, Trivedi et al. (2014)[7] showed that patients in certain race/ethnicity groups tend to experience differential, lower quality, or discriminatory care. Alternatively, patients with particular socioeconomic traits, such as lower education or language barriers, may require differentiated care (e.g., provision of lower literacy information) that they do not receive.

These potential pathways are complex to distinguish and have different implications on our decision to use them in the risk model or not. In general, we believe that certain demographic and physiologic factors, such as sex, age, and clinical symptoms at the start of care, are important for the model, while others, such as race, ethnicity, income (e.g., proxied by primary source of payer), and living environment, should be separately evaluated after a baseline model is developed. Our rationale is broadly twofold. First, using race/ethnicity as an example, there is no theoretical reason to believe that African Americans are more (or less) likely to experience AKI than Whites conditional on being acutely ill and hospitalized. Further, recent research and scientific advancement have removed race from the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation for the eGFR calculation and hence, it would appear logically inconsistent for the measure to use a race-neutral eGFR equation while adjusting for race in outcome measurement. Second, for factors related to SDoH (e.g., unstable living environment), we believe that it should be within hospitals' control to practice best care and adjust care processes for those with disrupted lifestyle or risky behaviors affected by their living environment.

Based on the published literature summarized by KDIGO and the available data elements from the electronic health record, we focused on the following risk variables: sex, age, vital signs (temperature, heart rate, respiratory rate, oxygen saturation, and systolic blood pressure), index eGFR, and comorbidities at the start of care. Vital signs were conceptualized as markers of dehydration, volume depletion, acute respiratory failure, sepsis, and other severe acute conditions. Comorbidities of primary interest included diabetes, heart failure, cancer, chronic liver disease, chronic lung disease, and others mentioned above.

Mediating Factors: Several care processes and intermediate factors (or mediators) may also contribute to the occurrence of AKI. For example, suboptimal inpatient management of blood pressure or heart failure may lead to fluctuations in renal perfusion, which in turn lead to transient increases in serum creatinine, meeting criteria for AKI. This pathway exemplifies a process of care that could be modified by health care providers, and thus should not be included in the risk-adjustment model.

Hospital length of stay (LOS) is a mediating factor that warranted careful examination. On the one hand, longer LOS can be reflective of a patient's underlying health status (i.e., severity of illness) that is not fully captured by vital signs or comorbidities. On the other hand, ineffective hospital care processes can extend hospital days beyond what was expected or normative for the patient's condition, especially when potentially preventable non-renal complications (e.g., hospital-acquired infections) arise. To examine whether LOS is a patient-level or a hospital-level risk factor in the context of AKI, we turned to empirical analyses. Specifically, we decomposed LOS for every denominator-qualifying encounter as follows:

Let LOS_{ij} be the LOS for encounter i at hospital j and LOS_j be the average LOS at hospital j across all denominator-qualifying encounters in the measurement period (i.e., calendar year 2020). Then LOS_{ij} can be re-stated as $(LOS_{ij} - LOS_j) + LOS_j$. We then included $(LOS_{ij} - LOS_j)$ and LOS_j in the regression at the same time, relating the probability of AKI to patient demographic characteristics, vital signs at the encounter start, index eGFR, and comorbidities. This analysis allowed us to assess the independent effect of two classifying groups simultaneously: 1) patient LOS on the probability of experiencing AKI when seen at a typical hospital and 2) hospitals with longer or shorter patient LOS on the rate of AKI for a typical patient.

The streamlined conceptual model helped us identify the following risk factors for testing in the development of baseline risk model: sex, age, vital signs, comorbidities, the index eGFR, and hospital length of stay (LOS).

We used several methods to determine risk factors' functional forms in the model. For indicators with only two values/categories, the form in which they exist in the model is straightforward. For continuous variables, such as age, vital signs, index eGFR, and LOS, we first ran locally weighted bivariate regressions (or lowess smoothing), relating the probability of AKI to each factor across its observed distribution and examining the potential empirical cutoff points for the factor. We then tabulated the frequency of denominator and numerator events for the empirically revealed cutoffs to gauge if further aggregation is needed to circumvent the small cell size issue. Third, when the first two exercises did not lend a clear support to the factor's functional form, we ran standard logit regressions, comparing one functional form against its alternative while controlling for the other risk factors in the model. This latter analysis allowed us to examine if a particular form is more prominently featured in data. Fourth, we compared model's goodness-of-fit between

functional forms on the basis of deviance and deviance ratio. For logit, the deviance and deviance ratio can be expressed as: $D = -\frac{2}{N_5}\sum_{i=1}^{N}[\tilde{y}Xb_i + \ln\{1 + rcp(Xb_i)\}]$

$$D^2 = \frac{D_{null} - D}{D_{null}}$$

where $\tilde{y_i} = 1\ if\ y_i > 0$ and 0 otherwise, matrix x subsumes variables included in the logistic regression, vector $b_i$denotes coefficient estimates calculated using data points *solely* from the training set (70% of the full sample). Given our interest in comparing risk factor's functional form, D and D^2 were calculated using data points from the validation sample (15% of the full sample). Data points in the training set and validation set are mutually exclusive. If D indicates the deviance calculated when all risk factors were in the model, then D_{null} is the deviance calculated when only a constant term is included in the logit. Hence, $D_{null} = -\frac{2}{N_5}\sum_{i=1}^{N}[\tilde{y_i}\ln\tilde{y} + (1-\tilde{y_i})\ln(1-\tilde{y})]$

where $\bar{y} = \frac{1}{N_s}\sum_{i=1}^{N}\tilde{y_i}$ denotes the proportion of denominator encounters where patients had suffered AKI. The smaller the D (or the larger the D^2) the better the model's out-of-sample fit. Lastly, we brought in clinical knowledge from the published literature to fine-tune the risk factors' functional forms.

Upon determining the risk factors' functional forms for testing in the baseline risk model build, we then examined if their presence in the risk model bears empirically relevance by fitting logistic lassos using three selection methods described in section 2b.20 and data points from the training set. The penalized objective function can be generalized as: $Qmin = \frac{1}{N}\sum_{i=1}^{N}f(y_i, \beta_0 + x_i\beta') + \lambda\sum_{j=1}^{p}k_j|\beta_j|$ where N is the total number of observations in training set, $f(\cdot)$ is the likelihood contribution for the logit model, i.e., $f(\beta_0 + x_i\beta') = -y_i(\beta_0 + x_i\beta') + \ln\{1 + exp(\beta_0 + x_i\beta'\}$, β_0 is the intercept, x_i is the $1 \times p$ vector of covariates, β is the $1 \times p$ vector of coefficients, λ is the lasso penalty parameter that is strictly non-negative, and k_j are coefficient-level weights.

The first selection method we used is the 10-fold CV, where the optimal λ, or λ^*, is chosen by the minimum of CV function (i.e., out-of-sample prediction error) across all CV functions computed for every λ in the λ grid. The 10-fold CV is widely used in the field when the goal is prediction and is a popular approach in developing risk adjustment methodology for performance measures.

The second selection method we employed is a variant of the 10-fold CV, in which λ^* is chosen not by the minimum of CV function (CV-min) but by the largest λ for which CV function lies within one standard error of CV-min. Hastie, Tibshirani, and Wainwright (2015)[8] promoted this approach, with the goal to emphasize parsimony without sacrificing model's generalizability. Although some authors[9] have argued that the one standard error (1SE) rule can have comparable model performance while alleviating the over-selection tendency of the traditional CV-min approach, its comparative advantage was not obvious in our data, given relatively few potential covariates.

Our third selection method picks λ^* based on the minimum of the BIC function (BIC-min). Comparing to the 10-fold CV-min approach, this method is less computationally intensive, also avoids the over-selection issue, and can produce a more parsimonious model. Zhang, Li, and Tsai (2010)[10] showed that, under certain conditions, λ^* selected by BIC-min will yield a set of covariates close to the true set that is needed to achieve optimal model prediction. The BIC function is defined as: $BIC = -2\log L(y, \beta_0 + x_i\beta') + df \cdot \log(N)$ where $\log L(y, \beta_0 + x_i\beta')$ is the log-likelihood function, df is the number of nonzero coefficients, and the rest parameters are defined as above.

Following the initial model fit on the training set, we calculated deviance and deviance ratio using post-selection coefficient estimates on the training set and separately on the validation set to evaluate model's in-sample and out-of-sample goodness-of-fit. Post-selection coefficient estimates were calculated by taking the covariates selected by a given lasso run (CV-min, 1SE rule, or BIC-min) and re-estimating the coefficients using an unpenalized estimator, i.e., logistic regression. We also followed Nattino et al. (2017)[11] by graphing calibration belts based on features selected under each method and on the training set and validation set, respectively. The calibration belt is an advancement of the conventional Hosmer-Lemeshow plot, as the latter has the limitation of undue sensitivity to the choice of bins (e.g., 10 bins split in deciles) and extreme fluctuations in the observed-to-expected ratios in bins with few harm events. Lastly, we

examined selected features across the three selection methods through the lens of clinical soundness, taking into account the need for face validity.

References:

1. Annual data report. USRDS. https://adr.usrds.org/2020/chronic-kidney-disease/5-acute-kidney-injury.
2. Heung M, Chawla LS. Predicting progression to chronic kidney disease after recovery from acute kidney injury. *Curr Opin Nephrol Hypertens*. 2012;21(6):628-634.
3. Hoste EA, Schurgers M. Epidemiology of acute kidney injury: how big is the problem?. *Crit Care Med*. 2008;36(4 Suppl):S146-S151.
4. Kidney Disease: Improving Global Outcomes (KDIGO). (2012). KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease (Vol. 2).
5. Jha, A. K., Orav, E. J., & Epstein, A. M. (2011). Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and Medicaid patients. *Health Affairs*, *30*(10), 1904-1911.
6. Reames, B. N., Birkmeyer, N. J., Dimick, J. B., & Ghaferi, A. A. (2014). Socioeconomic disparities in mortality after cancer surgery: failure to rescue. *JAMA Surgery*, *149*(5), 475-481.
7. Trivedi, A. N., Nsa, W., Hausmann, L. R., Lee, J. S., Ma, A., Bratzler, D. W., ... & Fine, M. J. (2014). Quality and equity of care in US hospitals. New England Journal of Medicine, 371(24), 2298-2308.
8. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
9. Chen, Y., & Yang, Y. (2021). The One Standard Error Rule for Model Selection: Does It Work?. *Stats*, *4*(4), 868-892.
10. Zhang, Y., Li, R., & Tsai, C. L. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*, *105*(489), 312-323.
11. Nattino, G., Lemeshow, S., Phillips, G., Finazzi, S., & Bertolini, G. (2017). Assessing the calibration of dichotomous outcome models with the calibration belt. *The Stata Journal*, *17*(4), 1003-1014

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

As discussed in section 2b.23 we ran decomposition analysis to determine whether LOS is a patient-level risk factor or a hospital-level effect in the context of AKI. If LOS is mostly a patient-level effect, then it should be treated as a risk factor to capture patients' underlying severity of illness that is not fully capturable by vital signs or comorbidities. On the other hand, if LOS is mostly a hospital-level effect, then the risk model should not adjust for it. We found that the encounter-level (patient-level) effect is much more prominent than the hospital-level effect, with the former being statistically significant at the 0.1% level and the latter at the 10% level. More importantly, the magnitude of encounter-level effect was almost double that of hospital-level effect, on average. **Exhibit 5** below summarizes our empirical analysis by showing the ratio of encounter-level effect to hospital-level effect (red circles and purple triangles) and the range of ratio (left panel) or the 95% confidence interval of ratio (right panel) across the 20 sites and stratified by EHR systems.

Exhibit 5. Decomposition Analysis on The Effect of Hospital Length of Stay

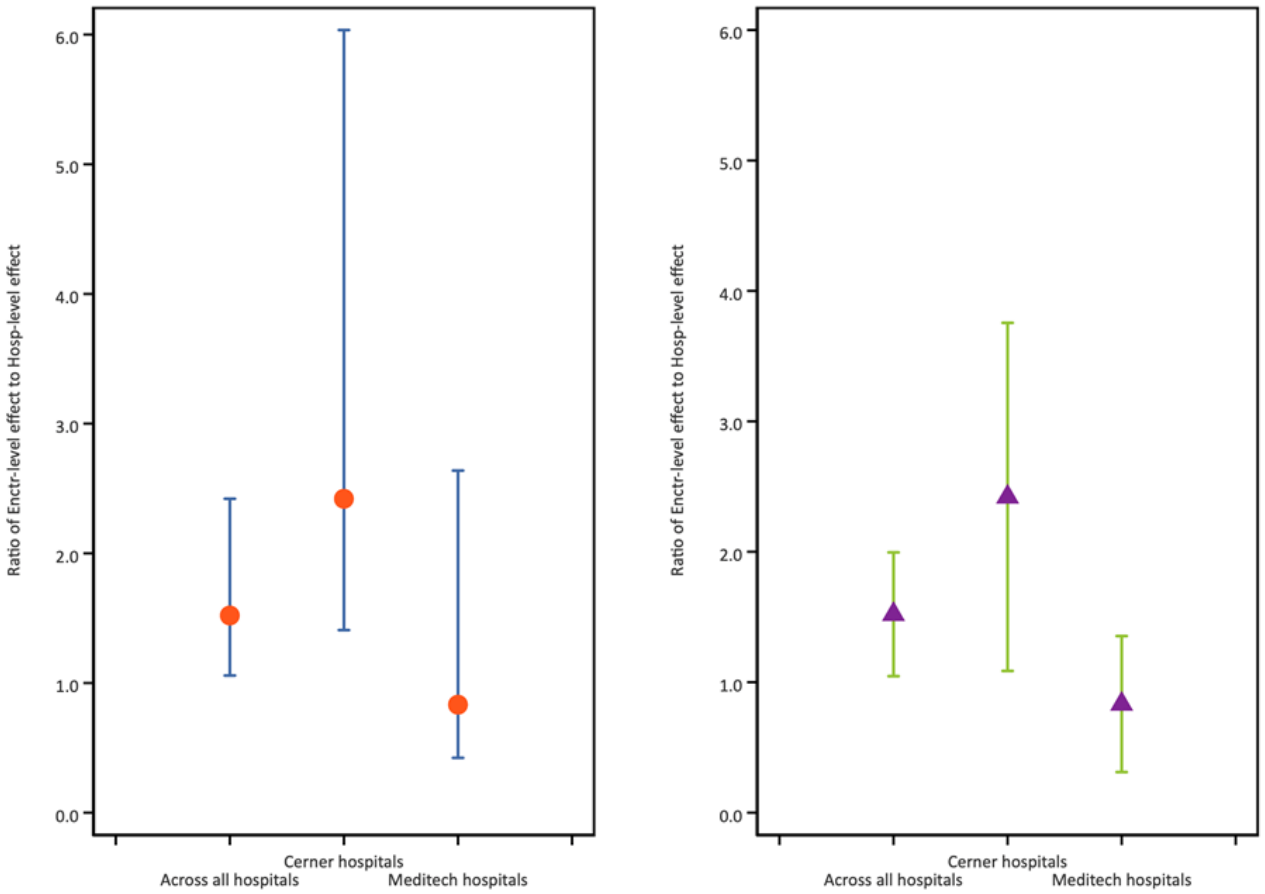


Exhibit 5 shows the ratio of encounter-level effect to hospital level effect and the range of the ratio on the left panel and the 95% confidence interval on the right panel.

Note: red circles or purple triangles denote the ratio of encounter-level effect to hospital-level effect. Vertical capped bars on the left panel show the range of such ratio, calculated as the 95% upper bound of encounter-level effect to the 95% lower bound of hospital-level effect and 95% lower bound of encounter-level effect to the 95% upper bound of hospital-level effect. Vertical capped bars on the right panel show the 95% confidence interval of such ratio.

Table 25 below shows the numerator and denominator prevalences and AKI rate, stratified by patient sex-age category interactions and with data points from the full sample. As discussed in section 2b.23, we used different methods to determine the functional form "age" should enter the risk model, and our analyses suggested grouping age into four bins (18-34, 35-54, 55-74, and 75 or older) and then interacting with patient sex.

Table 25. Numerator and Denominator Prevalence and AKI Rate; Four Age Categories by Patient Sex

Risk Factors	Numerator Count (N)	Denominator Count (N)	Measure Rate
Age 18-34 × Female	23	2,482	0.93%
Age 35-54 × Female	95	6,533	1.45%
Age 55-74 × Female	176	11,479	1.53%
Age 75+ × Female	100	7,380	1.36%
Age 18-34 × Male	35	3,023	1.16%
Age 35-54 × Male	117	7,963	1.47%
Age 55-74 × Male	248	13,887	1.79%

Risk Factors	Numerator Count (N)	Denominator Count (N)	Measure Rate
Age 75+ × Male	102	6,189	1.65%

Tables 27 to 31 below show the numerator and denominator prevalences and AKI rates for each of the five vital signs that are available to testing and captured at the start of care. Lowess smoothing pointed out strong non-linearity and empirical analyses and clinical knowledge allowed us to group each into mutually exclusive bins.

- **Temperature**: Category 1 (low abnormal): < 36°; Category 2 (normal): 36° – 38.4°; and Category 3 (high abnormal): > 38.4°
- **Heart Rate**: Category 1 (low abnormal): < 70 beats per minute (bpm); Category 2 (normal): 70 – 109 (bpm); and Category 3 (high abnormal): > 109 (bpm)
- **Respiratory Rate**: Category 1 (low abnormal): < 12 breaths per minute; Category 2 (normal): 12 – 24 breaths per minute; and Category 3 (high abnormal): > 24 breaths per minute
- **Oxygen saturation**: Category 1 (indicative of hypoxemia): < 90% and Category 2 (normal): 90 – 100%
- **Systolic blood pressure**: Category 1: < 110 (mm Hg); Category 2: 110 – 160 (mm Hg); and Category 3: > 160 (mm Hg)

Table 26. Numerator and Denominator Counts and AKI Rate, Temperature

Risk Factor: Temperature	Numerator Count (N)	Denominator Count (N)	AKI Rate
Low abnormal temperature	47	1,885	2.49%
Normal temperature	803	54,361	1.48%
High abnormal temperature	46	2,690	1.71%

Table 27. Numerator and Denominator Counts and AKI Rate, Heart Rate

Risk Factor: Heart Rate	Numerator Count (N)	Denominator Count (N)	AKI Rate
Low abnormal heart rate	106	7,772	1.36%
Normal heart rate	536	37,383	1.43%
High abnormal heart rate	254	13,781	1.84%

Table 28. Numerator and Denominator Counts and AKI Rate, Respiratory Rate

Risk Factor: Respiratory Rate	Numerator Count (N)	Denominator Count (N)	AKI Rate
Low abnormal respiratory rate	15	538	2.79%
Normal respiratory rate	719	52,435	1.37%
High abnormal respiratory rate	162	5,963	2.72%

Table 29. Numerator and Denominator Counts and AKI Rate, Oxygen Saturation

Risk Factor: Oxygen Saturation	Numerator Count (N)	Denominator Count (N)	AKI Rate
< 90%	3,347	74	2.16%
90 – 100%	54,693	822	1.48%

Table 30. Numerator and Denominator Counts and AKI Rate, Systolic Blood Pressure

Risk Factor: Systolic Blood Pressure	Numerator Count (N)	Denominator Count (N)	AKI Rate
< 110 (mm Hg)	154	8,110	1.90%
110 – 160 (mm Hg)	555	40,096	1.38%
> 160 (mm Hg)	187	10,730	1.74%

We compared two functional forms of the index eGFR, with the first categorizing index eGFR into two mutually exclusive bins (> 60 but < 90 mL/min; ≥ 90 mL/min) and the second assuming linearity (lowess smoothing did not offer strong support for non-linearity). As discussed in section 2b.23, we ran logit regressions, comparing the statistical property between the two functional forms by simultaneously controlling for the other risk factors in the model. **Table 31** compares the predictive power of the two functional forms of index eGFR. Model performance (both in-sample and out-of-sample) was slightly better when index eGFR was entered in linear form.

Table 31. Deviance and Deviance Ratio; Index eGFR

Functional form	Method	Sample	Deviance	Deviance ratio	Observation
Linear	Logit	Training	0.1529	0.0387	41,257
*	*	Validation	0.1513	0.0311	8,839
Two groups	Logit	Training	0.1529	0.0383	41,257
*	*	Validation	0.1514	0.0307	8,839

Notes: logit regressions controlled for the other risk factors. Training set accounts for 70% of the full sample and validation set accounts for 15% of the full sample. Data points in these two sets are mutually exclusive. Deviance and deviance ratio calculated using data points from the validation set were based on the coefficient estimates derived from the training set.

*Cells intentionally left empty.

We ran similar logit regressions to compare the functional forms of patient LOS, with the first assuming linearity and the second grouping LOS into four mutually exclusive bins (0-7; 8-14; 15-21; 22-30). **Table 32** compares the predictive power of the two functional forms of LOS. Here, categorizing LOS into bins led to better model performance.

Table 32. Deviance and Deviance Ratio; Encounter LOS

Functional form	Method	Sample	Deviance	Deviance ratio	Observation
Linear	Logit	Training	0.1270	0.2015	41,257
*	*	Validation	0.1272	0.1595	8,839
Four groups	Logit	Training	0.1254	0.2114	41,257
*	*	Validation	0.1237	0.1830	8,839

Notes: logit regressions controlled for the other risk factors. Training set accounts for 70% of the full sample and validation set accounts for 15% of the full sample. Data points in these two sets are mutually exclusive. Deviance and deviance ratio calculated using data points from the validation set were based on the coefficient estimates derived from the training set. *Cells intentionally left empty.

Given the findings above, we split LOS into four bins for model development. **Table 33** below shows the numerator and denominator prevalences and AKI rates for each of the four LOS categories.

Table 33. Numerator and Denominator Counts and AKI Rate, Patient LOS

Risk Factor: LOS	Numerator Count (N)	Denominator Count (N)	AKI Rate
0 – 7d	144	43,492	0.33%
8 – 14d	342	11,959	2.86%
15 – 21d	193	2,196	8.79%
22 – 30d	217	1,289	16.83%

Notes: measure caps patient LOS at the 30th day since the encounter start.

We captured patients' comorbidities using the AHRQ Elixhauser Comorbidity Software Refined for ICD-10-CM (version 2022.1). The AHRQ Elixhauser software is coded using publicly available Healthcare Cost and Utilization Project (HCUP) software, annually updated, and extensively validated. For the development of the baseline risk model, we grouped together some related comorbidities due to small sample size concerns. Specifically,

- Diabetes with chronic complications + Diabetes without chronic complications $\Longrightarrow$ Diabetes
- Hypertension, complicated + Hypertension, uncomplicated $\Longrightarrow$ Hypertension
- Hypothyroidism + Other thyroid disorders $\Longrightarrow$ Thyroid disorders
- Leukemia + Lymphoma + Metastatic cancer $\Longrightarrow$ Cancer (LLM)

Table 34 below shows the numerator and denominator prevalences and AKI rates for each of the evaluated comorbidities.

Table 34. Numerator and Denominator Counts and AKI Rate, Elixhauser Comorbidities

Risk Factors: Comorbidities	Numerator Count (N)	Denominator Count (N)	Measure Rate
Acquired immune deficiency syndrome - No	882	58,581	1.51%
Acquired immune deficiency syndrome - Yes	14	355	3.94%
Alcohol abuse - No	797	53,161	1.50%
Alcohol abuse - Yes	99	5,775	1.71%
Autoimmune conditions - No	855	56,654	1.51%
Autoimmune conditions - Yes	41	2,282	1.80%
Cancer (LLM) - No	788	54,860	1.44%
Cancer (LLM) - Yes	108	4,076	2.65%
Cancer (Other) - No	768	52,874	1.45%
Cancer (Other) - Yes	128	6,062	2.11%
Dementia - No	812	54,551	1.49%
Dementia - Yes	84	4,385	1.92%
Depression - No	748	49,730	1.50%
Depression - Yes	148	9,206	1.61%
Diabetes - No	522	40,380	1.29%

Risk Factors: Comorbidities	Numerator Count (N)	Denominator Count (N)	Measure Rate
Diabetes - Yes	374	18,556	2.02%
Drug abuse - No	814	53,157	1.53%
Drug abuse - Yes	82	5,779	1.42%
Hypertension - No	243	22,046	1.10%
Hypertension - Yes	653	36,890	1.77%
Chronic pulmonary disease - No	654	43,441	1.51%
Chronic pulmonary disease - Yes	242	15,495	1.56%
Obese - No	595	45,520	1.31%
Obese - Yes	301	13,416	2.24%
Peripheral vascular disease - No	819	54,721	1.50%
Peripheral vascular disease - Yes	77	4,215	1.83%
Thyroid disorders - No	777	50,119	1.55%
Thyroid disorders - Yes	119	8,817	1.35%
Heart failure - No	632	48,991	1.29%
Heart failure - Yes	264	9,945	2.65%

Notes: Comorbidities were identified using AHRQ Elixhauser Comorbidity Software Refined for ICD-10-CM (version 2022.1) and patients' ICD-10-CM diagnoses only. POA information is not necessary for identification except for heart failure, where we have valid POA data to measure it. Cancer (LLM) includes leukemia, lymphoma, and metastatic cancer. Cancer (Other) denotes solid tumor without metastasis, malignant. Diabetes includes diabetes with and without complications. Hypertension includes hypertension with and without complications. Thyroid disorders include hypothyroidism and other thyroid disorders. Due to the sample size constraint, grouping is needed to avoid zero numerator count scenarios.

Table 35 below presents selected features for the three lasso estimations on the training data set, as well as their post-selection coefficient estimates obtained by fitting a logistic regression with the selected variables.

Table 35. Selected Features and Postselection Coefficient Estimates

*	CV-min	1SE rule	BIC-min
Sex-by-Age groups	*	*	*
Female×Age (18-34) - Reference group	–	–	–
Female×Age (35-54)	0.201	×	×
Female×Age (55-74)	×	×	×
Female×Age (75 and older)	×	×	×
Male×Age (18-34)	0.306	×	×
Male×Age (35-54)	×	×	×
Male×Age (55-74)	×	×	×

*	CV-min	1SE rule	BIC-min
Male×Age (75 and older)	×	×	×
Vital Signs	*	*	*
Abnormally low temperature	0.435	×	×
Normal temperature - Reference group	–	–	–
Abnormally high temperature	×	×	×
Abnormally low heart rate	×	×	×
Normal heart rate - Reference group	–	–	–
Abnormally high heart rate	0.216	×	×
Abnormally low respiratory rate	0.576	×	×
Normal respiratory rate - Reference group	–	–	–
Abnormally high respiratory rate	0.375	×	0.360
SBP < 110 (mm Hg)	0.171	×	×
SBP ≥110 & ≤ 160 (mm Hg) - Reference group	–	–	–
SBP > 160 (mm Hg)	0.250	×	×
Oxygen saturation ≥ 90% - Reference group	–	–	–
Oxygen saturation < 90%	-0.206	×	×
Index eGFR (mL/min)	-0.010	×	-0.009
Elixhauser Comorbidities	*	*	*
Acquired immune deficiency syndrome	0.432	×	×
Alcohol abuse	×	×	×
Autoimmune conditions	0.341	×	×
Cancer LLM	0.507	×	0.567
Cancer (Other)	0.162	×	×
Dementia	0.170	×	×
Depression	×	×	×
Diabetes	0.137	×	0.146
Drug abuse	-0.129	×	×
Hypertension	0.249	×	0.212
Chronic pulmonary disease	-0.188	×	×
Obesity	0.380	×	0.361
Peripheral vascular disease	-0.113	×	×
Thyroid disorders	-0.159	×	×

*	CV-min	1SE rule	BIC-min
Heart failure	0.569	×	0.499
Encounter LOS	*	*	*
>0d & ≤ 7d - Reference group	–	–	–
>7d & ≤14d	2.083	2.154	2.082
>14d & ≤21d	3.278	3.339	3.260
>21d & ≤30d	4.044	4.094	4.027

Notes: The CV-min approach picks λ^* based on the minimum of CV function and the 1SE rule approach picks λ^* as the largest λ for which the CV function lies within a standard error of the minimum of CV function. Both the CV-min and 1SE rule are based on the 10-fold data partition. The BIC-min approach picks λ^* based on the minimum of BIC function. "–" indicates that the variable/feature is omitted from the model and "×" indicates that the variable/feature is not selected by the lasso. Constant is included in the model but omitted from display.

*Cells intentionally left empty.

Table 36 shows the deviance and deviance ratio under the above three selection methods and stratified by dataset (i.e., in-sample or out-of-sample). The two metrics help to assess model fit with different sets of features included.

Table 36. Model's In-sample and Out-of-sample Goodness of Fit Based on Post-selection Coefficients

Method	Sample	No. of features selected	Deviance	Deviance ratio	Observation
CV-min	Training	26	0.1251	0.2132	41,257
CV-min	Validation	26	0.1240	0.1809	8,839
1SE rule	Training	3	0.1290	0.1891	41,257
1SE rule	Validation	3	0.1251	0.1735	8,839
BIC-min	Training	10	0.1260	0.2075	41,257
BIC-min	Validation	10	0.1243	0.1789	8,839

Notes: The CV-min approach picks λ^*based on the minimum of CV function and the 1SE rule approach picks λ^* as the largest λ for which the CV function lies within a standard error of the minimum of CV function. Both CV-min and 1SE rule are based on the 10-fold data partition. The BIC-min approach picks λ^* based on the minimum of BIC function. S.E = standard error. No. of features selected denotes the minimum number of nonzero coefficients that satisfy the objective function (CV-min, 1SE rule, or BIC-min). Models were fit on the training set only.

Both the 1SE rule and BIC-min approaches suggested a more parsimonious model than did the CV-min approach. However, the deviance ratio based on post-selection coefficients from the validation set was 5% and 1% smaller, respectively, than that based on the CV-min approach. Since a larger out-of-sample deviance ratio suggests better model generalizability, we rejected the 1SE rule for this application.

We then examined selected features (**Table 35**) through the lens of clinical soundness, taking into account the need for face validity. We believe that sex-by-age group interactions are essential for the risk model even if only two of them were deemed necessary for prediction when each combination was treated as an independent feature by lasso. Omitting certain sex-by-age groups would complicate interpretation of the risk model.

Vital signs recorded at the start of care (i.e., temperature, heart rate, respiratory rate, and systolic blood pressure) are useful components in the risk model because they may reflect major physiologic derangements that other observable factors do not, and they can proxy for other risk factors that are not available in structured fields. Given the findings for

temperature and heart rate, we further collapsed the original three groups into two groups and used them in the risk model. Due to the sample size constraint, we could only use hypoxemia (i.e., oxygen saturation less than 90%) as a marker of cardiopulmonary illness. However, the empirical impact of hypoxemia was actually favorable, holding everything else constant, due to collinearity with other features already selected by the model. As a result, we excluded oxygen saturation from the risk model.

Patients' baseline kidney function measured by index eGFR values is a strong and robust predictor for AKI and hence, we included it in the risk model.

The five Elixhauser comorbidities (cancer [llm], diabetes, hypertension, obesity, and heart failure) selected by CV-min and BIC-min are consistent with our expectation since all are well supported by the literature. For example, patients receiving nephrotoxic chemotherapy during hospitalization are at high risk for tumor lysis syndrome, and renal disease can result from drug treatment.[1] According to the Centers for Disease Control and Prevention (CDC), diabetes and hypertension are the leading causes of renal disease.[2] As described above, heart failure and reduced left ventricular ejection fraction are known to be associated with falling renal function.[3]

For the remaining comorbidities (AIDS, autoimmune conditions, cancer [other], dementia, drug abuse, chronic pulmonary disease, peripheral vascular disease, and thyroid disorders) selected only by the CV-min approach, clinical knowledge and empirical findings suggest that their presence in the risk model may not be necessary. For example, AIDS is relatively uncommon among hospitalized patients. Some antiretroviral medications may be nephrotoxic, and yet it is incumbent on clinicians to properly manage these medications to minimize nephrotoxicity.[4] Empirically, we observed implausible protective effects for drug abuse, chronic pulmonary disease, peripheral vascular disease, and thyroid disorders. As a result, we omitted these comorbidities from the risk model.

Lastly, encounter LOS is a robust predictor for AKI across lasso approaches, in line with clinical expectation that patients requiring lengthy hospital stays are at greater risk than other patients.

References:

1. Małyszko, J., Kozłowska, K., Kozłowski, L., & Małyszko, J. (2017). Nephrotoxicity of anticancer treatment. *Nephrology Dialysis Transplantation*, *32*(6), 924-936.
2. Centers for Disease Control and Prevention. *Chronic Kidney Disease in the United States, 2019*. Atlanta, GA: US Department of Health and Human Services. Centers for Disease Control and Prevention; 2019.
3. Silverberg, D., Wexler, D., Blum, M., Schwartz, D., & Iaina, A. (2004). The association between congestive heart failure and chronic renal disease. *Current opinion in nephrology and hypertension*, *13*(2), 163-170.
4. Kalyesubula, R., & Perazella, M. A. (2011). Nephrotoxicity of HAART. *AIDS research and treatment*, *2011.*

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

Our baseline risk model does not include social risk factors, such as race, ethnicity, primary source of payment, and variables reflecting patients' SDoH, which may serve as proxies for patients' underlying functional and physiologic status. For the theoretical reasons, please see section 2b.23.

To empirically assess if their presence in the model leads to stronger model performance, we augmented the set of risk factors determined above with patient race, ethnicity, primary source of payment, and an indicator for whether they experienced problems in housing and economic circumstances using the ICD-10-CM Z codes, and then compared the change in model performance, evaluated by AUROC and AURPC.[1] Problems/risk factors included in this Z code were

homelessness, inadequate housing, discord with neighbors, lodgers and landlord, problems related to living in residential institutions, lack of adequate food and safe drinking water, extreme poverty, low income, insufficient social insurance and welfare support. A total of nine categories supported by the Z codes (Z55-Z65) can be used to capture patients' SDoH, but sample size constraints limited our use of only one.

Table 37 below shows numerator and denominator prevalences and AKI rate for the social risk factors.

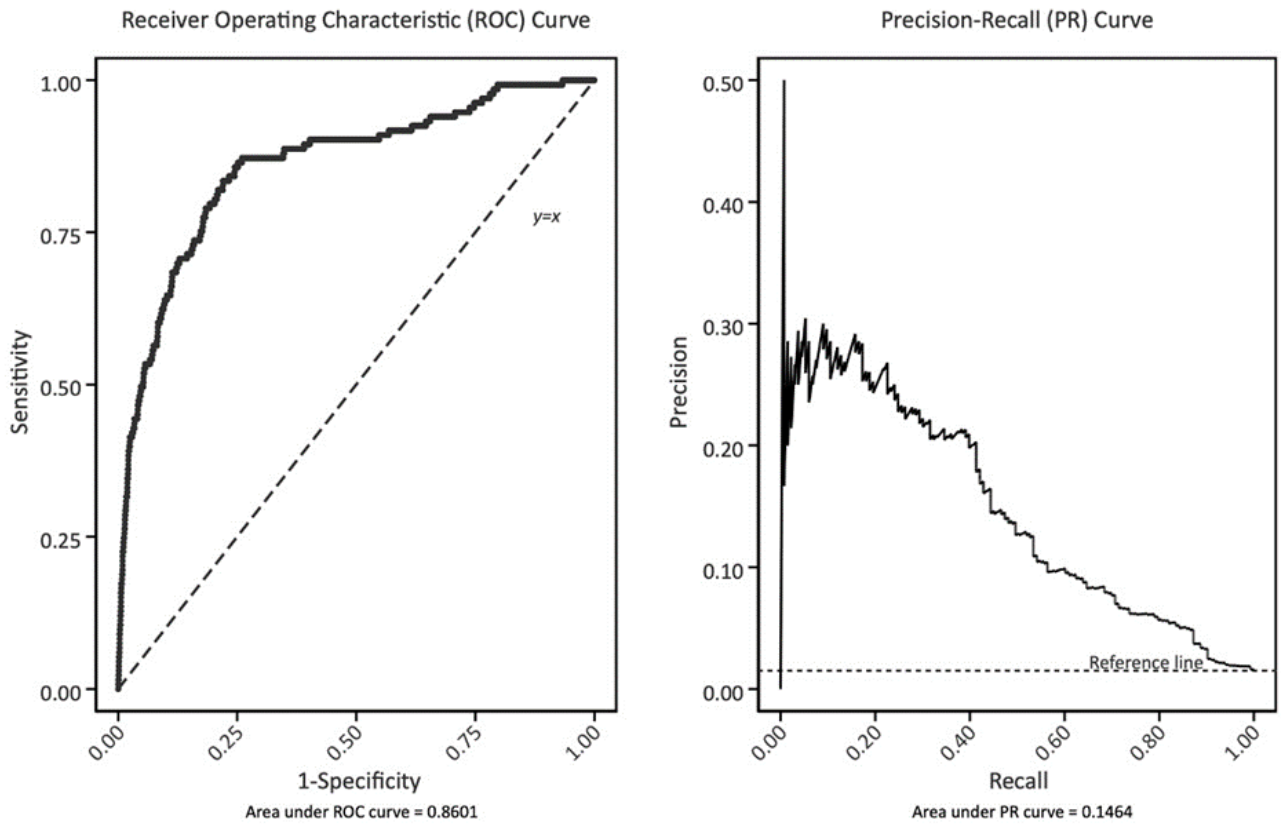
Table 37. Numerator and Denominator Counts and AKI Rate; Four Social Risk Factors

Social Risk Factors	Numerator Count (N)	Denominator Count (N)	AKI Rate
White	584	41,059	1.42%
Black	72	4,135	1.74%
Other	184	9,737	1.89%
Unknown	56	4,005	1.40%
Hispanic	174	10,552	1.65%
Non-Hispanic	647	43,117	1.50%
Unknown	75	5,267	1.42%
Medicare	180	10,972	1.64%
Medicaid	275	19,239	1.43%
Commercial	268	16,984	1.58%
Self-pay	92	8,186	1.12%
Other	25	2,084	1.20%
Unknown	56	1,471	3.81%
Problems related to housing and economic circumstances – No	876	56,891	1.54%
Problems related to housing and economic circumstances – Yes	20	2,045	0.98%

Note: Indicator "Problems related to housing and economic circumstances" coded based on ICD-10-CM Z59.

Exhibit 6 below shows ROC and PR curves on the holdout set using coefficient estimates (with the added social risk factors in the table above) derived from the training set. Comparing it to the graph included in section below (section 2b.26), we see no model performance enhancement by adding patient race, ethnicity, primary source of payment, and an indicator for housing and economic hardship.

Exhibit 6. ROC Curve and PR Curve; Holdout Set



Note: Both curves were generated on the holdout set using coefficient estimates derived from the training set. Social risk factors shown in **Table 37** were included in the logit regression.

Given the lack of clear theory supporting the relationship between social risk factors and AKI as well as our empirical findings, inclusion of these factors in the baseline risk model is not justified.

References:

1. *Resource on ICD-10-CM coding for Social Determinants of Health: AHA*. American Hospital Association. (n.d.). Retrieved February 6, 2022, from https://www.aha.org/dataset/2018-04-10-resource-icd-10-cm-coding-social-determinants-health

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

We assessed model adequacy (i.e., model's out-of-sample goodness-of-fit) using deviance, deviance ratio, and calibration belts suggested by Nattino et al. (2017).[1] Calibration belts are an advance over the conventional Hosmer-Lemeshow plot, as the latter has the limitation of undue sensitivity to the choice of bins (e.g., 10 bins split in deciles) and extreme fluctuations in the observed-to-expected ratios in bins with few harm events.

We used two metrics to evaluate model performance using data from the holdout set, which were not used for model training or parameter tuning and hence allow unbiased assessment of the model:

- Receiver operating characteristics (ROC) curve and the area under the curve (AUROC or C-statistic). The C-statistic is a well-known metric used in model performance evaluation, and has intuitive appeal as it represents a model's capacity to discriminate by consistently assigning higher risk estimates to randomly selected patients who experienced AKI than to patients who did not. Yet, the ROC curve and C-statistic tend to offer an inflated view of model performance when applied to "rare event" data.
- Precision-recall (PR) curve and the area under the curve (AUPRC). The PR curve and AUPRC are less sensitive to data imbalance or class imbalance than the AUROC.

Using the canonical confusion matrix (**Table 38** below), ROC curves plot recall (true-positive rate or sensitivity) as a function of the false-positive rate (or 1 – specificity) and PR curves plot precision (percent of cases predicted to be positive cases are truly positive cases) as a function of recall. Recall is defined as $\frac{TP}{P}$, precision is defined as $\frac{TP}{TP+FP}$, and false-positive rate is defined as $\frac{FP}{N}$.

Table 38. Confusion Matrix Defining TP, FP, Negatives, and Positives

*	*	Truth	Truth
*	*	Positive	Negative
Prediction	Positive	True positive (TP)	False positive (FP)
Prediction	Negative	False negative (FN)	True negative (TN)
Prediction	Total	Positive (P)	Negative (N)

Note: *Cells intentionally left empty.

References:

1. Nattino, G., Lemeshow, S., Phillips, G., Finazzi, S., & Bertolini, G. (2017). Assessing the calibration of dichotomous outcome models with the calibration belt. *The Stata Journal*, *17*(4), 1003-1014

[Response Ends]

2b.27. Provide risk model discrimination statistics.

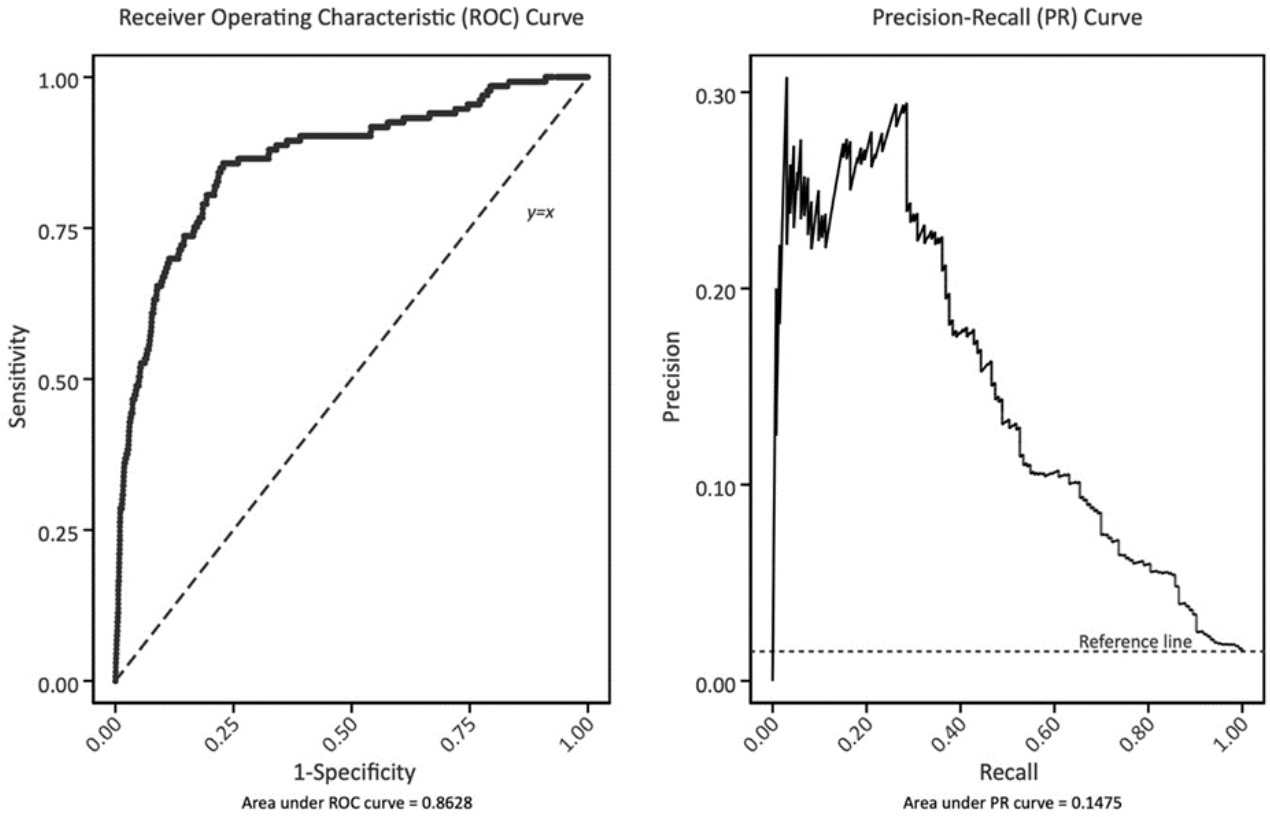
For example, provide c-statistics or R-squared values.

[Response Begins]

Exhibit 7 shows ROC and PR curves for the baseline risk model on the holdout set using coefficient estimates derived from the training set. AUROC (or C-statistic) and AUPRC are shown on the bottom of the graph, and the dashed lines (reference lines) indicate how a completely uninformative model would perform. For the ROC curve, this line corresponds to the 45-degree identity line. For the PR curve, it corresponds to the rate of AKI in the full dataset. ROC curve moves towards the upper left as performance improves, whereas the PR curve moves towards the upper right as performance improves.

The baseline risk model's C-statistic equals 0.863.

Exhibit 7. ROC Curve and PR Curve; AUROC and AUPRC; Holdout Set



[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

As shown in **Exhibit 8** in section 2b.29, the likelihood ratio chi-square statistic equals 1.18 with p-value equal to 0.554 in the validation (NOT model development) sample.

[Response Ends]

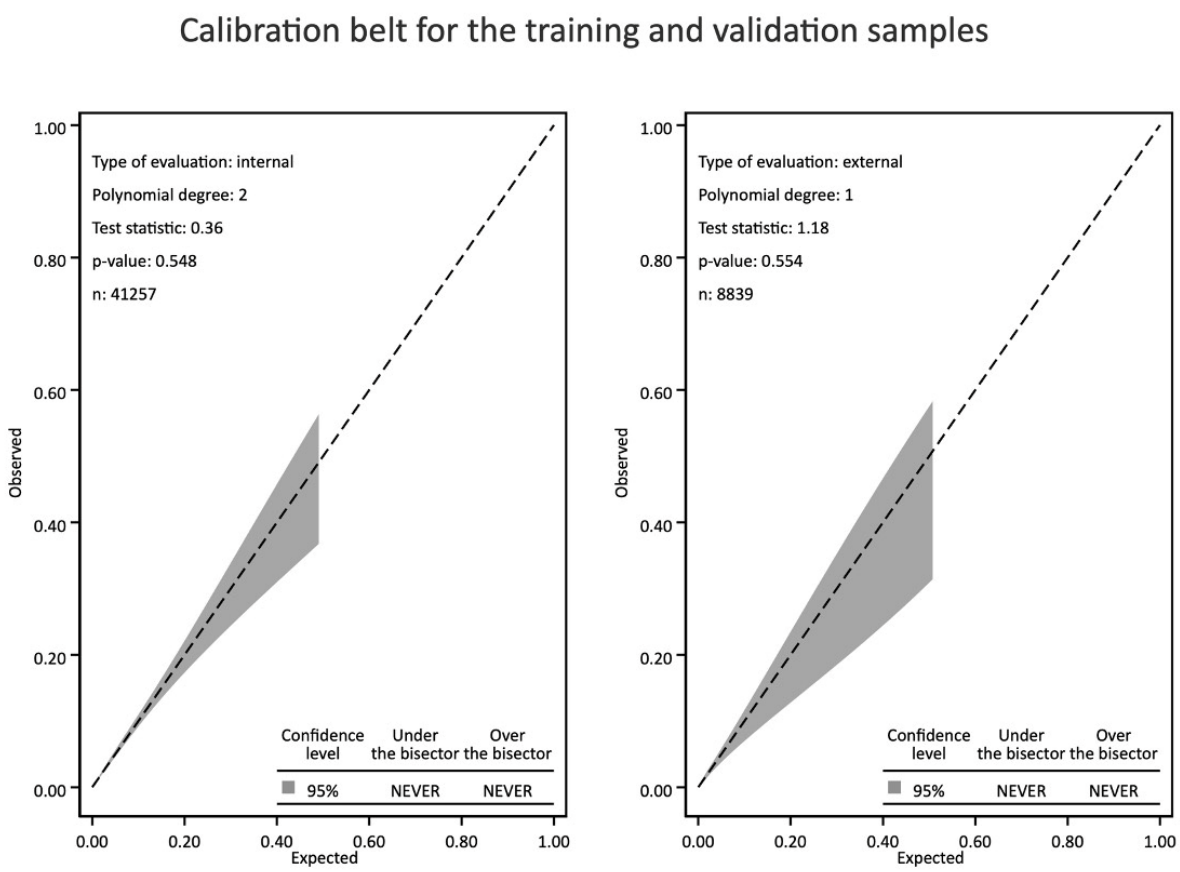
2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

Exhibit 8 shows calibration belts for the training set and validation set, respectively, using the determined risk factors and their functional forms. Test statistics (i.e., likelihood-ratio statistics and p-values) show that strong out-of-sample model calibration and goodness-of-fit cannot be rejected at any conventional levels. Similar conclusions can be drawn from the interpretation of plots. Specifically, the 95% calibration belts encompass the bisector over a decent range of the predicted probabilities. The predictions of the model do not significantly deviate from the observed rate in the validation sample and hence, model's external validation is satisfactory.

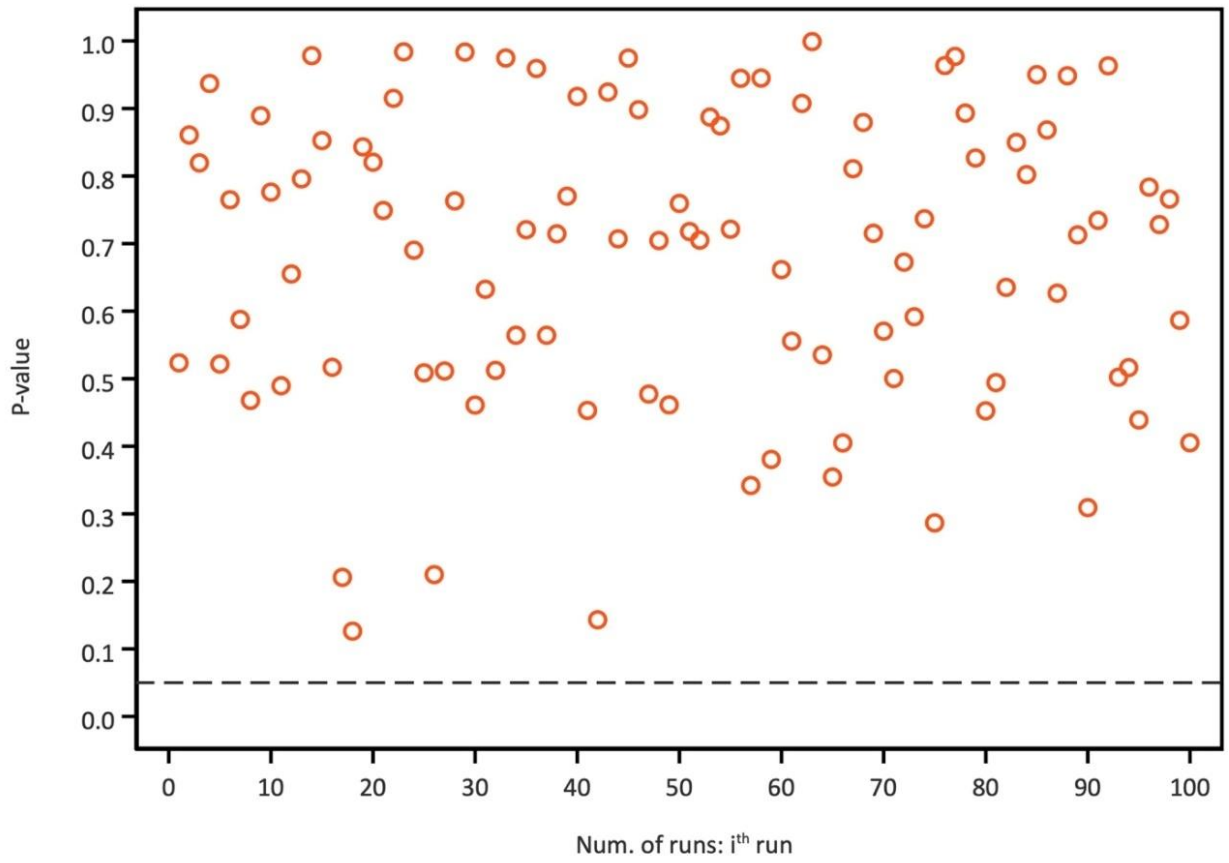
Exhibit 8. Calibration Belt Based on The Chosen Set of Risk Factors; Training Set and Validation Set Respectively



Notes: Risk factors are shown in **Table 37** above. Calibration belts were calculated using the chosen risk factors and data points only from the training set. Calibration belt on the left panel uses data points from the training set while calibration belt on the right panel uses data points from the validation set. The 45 dashed identity line indicates perfect model fit. "Under the bisector" indicates model overfit as predicted probabilities exceed observed probabilities. "Over the bisector" indicates model underfit as predicted probabilities below observed probabilities. A confidence band around the curve, namely, the calibration belt, reflects the statistical uncertainty about the estimate of the curve.

To confirm that the satisfactory model calibration was not driven by chance, we conducted a simple simulation exercise. By generating 100 random validation samples and 100 calibration belts in each of these samples using coefficient estimates derived from the training sample, we graphed the distribution of p-values. **Exhibit 9** below shows that the p-value never fell below 0.10, suggesting that the hypothesis of good calibration is not rejected at the 0.10 level.

Exhibit 9. Distribution of P-values From Calibration Belts In 100 Simulation Runs



Note: each circle denotes a p-value from the hypothesis test that evaluates the goodness of model calibration in one of the 100 simulated validation samples.

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

Not applicable.

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

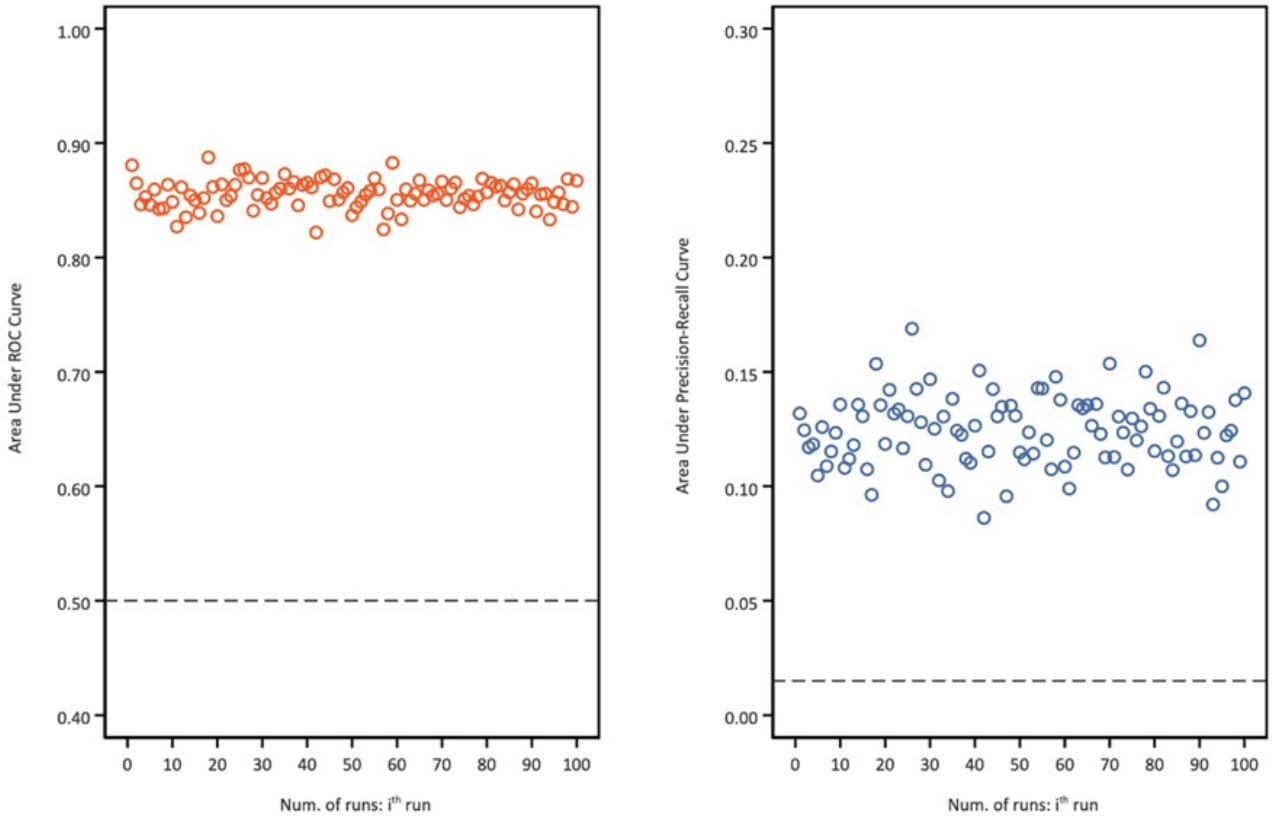
[Response Begins]

Exhibits 7 (section 2b.27) and **Exhibit 8** (section 2b.29) show strong model performance. In particular, C-statistic is larger than 0.8, which is a benchmark frequently cited for demonstrating excellent model performance. Similarly, AUPRC tells us that the baseline risk model is close to 15 times better than a random prediction. We caution against the

interpretation of AUPRC based on its absolute value, as by construction AUPRC is affected by the base rate and inversely related to data imbalance.

Recognizing that the test sample is relatively small, we performed a simulation exercise to gauge the extent to which satisfactory model performance may be driven by chance. By generating 100 random holdout sets and calculating 100 AUROC and AUPRC using coefficient estimates derived from the training set, we plot the distribution of AUROC and AUPRC. **Exhibit 10** below clearly shows that the odds of strong model performance occurring by chance are very small.

Exhibit 10. Distribution of AUROC and AUPRC In 100 Simulation Runs



Note: each circle in the first (left) figure denotes the AUROC calculated within a given simulated holdout set and each circle in the second (right) figure denotes the AUPRC calculated within a simulated holdout set. The horizontal dash lines serve as the benchmark, indicating that what the model performance would have been should it have zero predictive power.

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

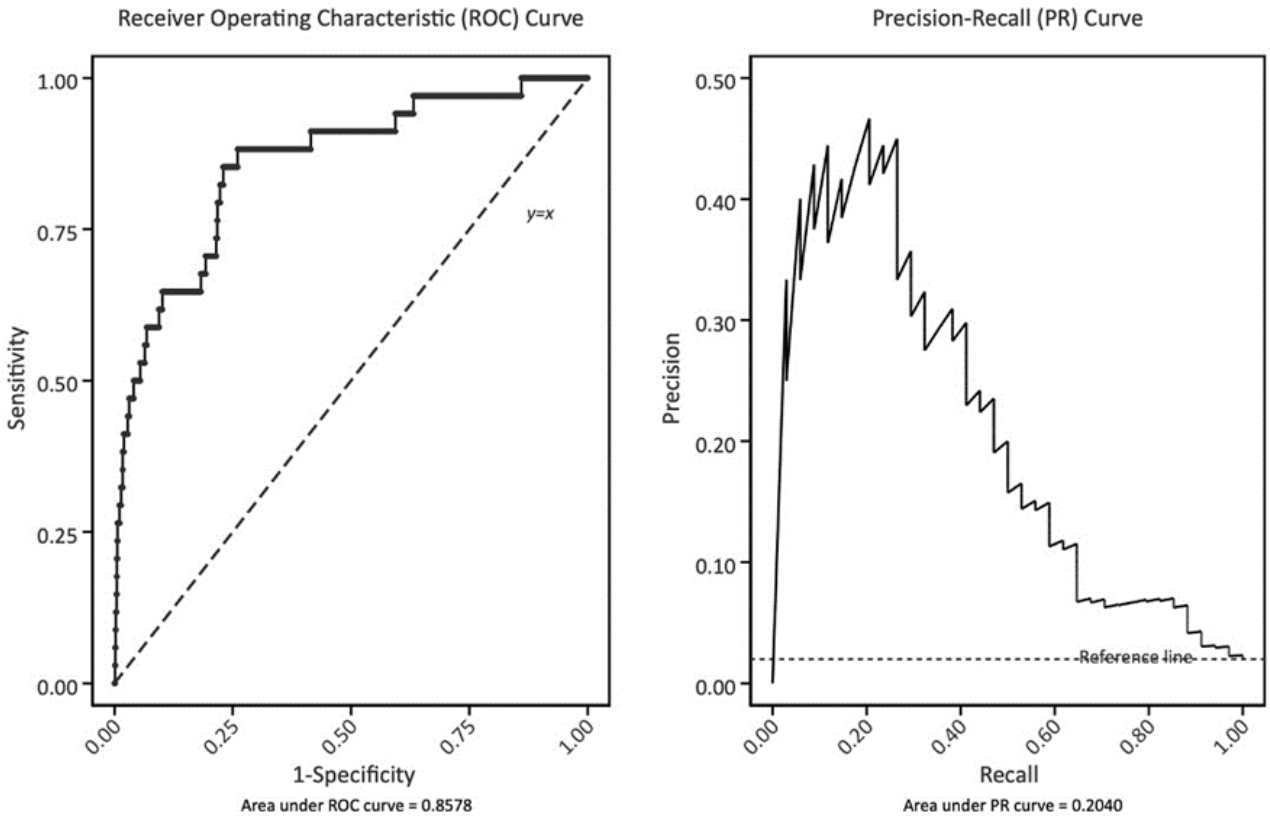
Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

We developed the baseline risk model using all-payer data, but one may be interested in model performance in sub-populations. **Exhibits 11** to **14** below show the ROC and PR curves and area under the curves for the Medicare, Medicaid, and Commercial populations, respectively. To graph these curves and compute their corresponding areas, we used the **already generated coefficient estimates** and data from the holdout set. We define Medicare population as those for

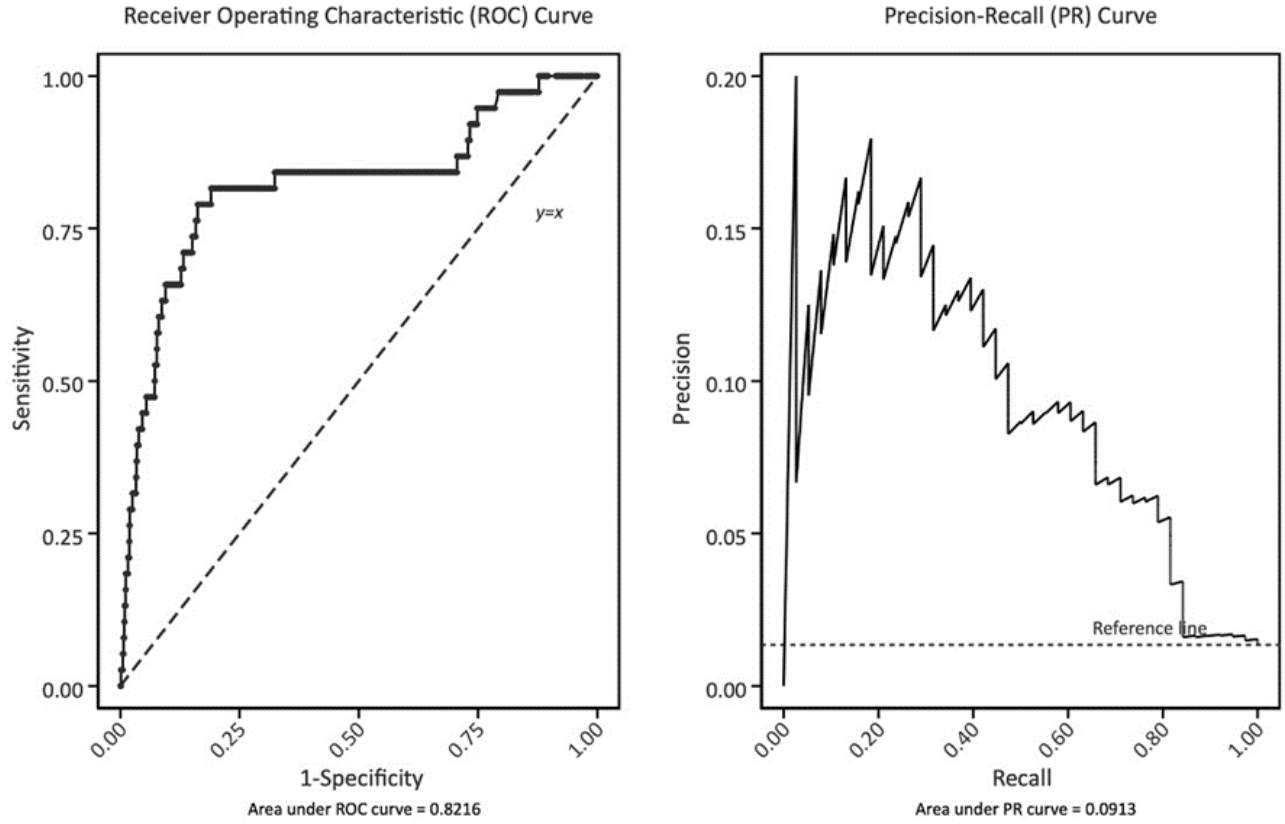
whom Medicare Fee-for-Service (FFS) is the primary source of payment. We define Medicaid population as those for whom Medicaid FFS or Medicaid managed care is the primary source of payment.[1] We define Commercial population as those for whom private insurance, including Medicare Advantage, is the primary source of payment. The three sub-populations accounted for 19%, 33%, and 29% of the measure denominator population, respectively. These analyses show that model performance is strong across all three payer categories.

Exhibit 11. ROC Curve and PR Curve; Holdout Set; Medicare Population



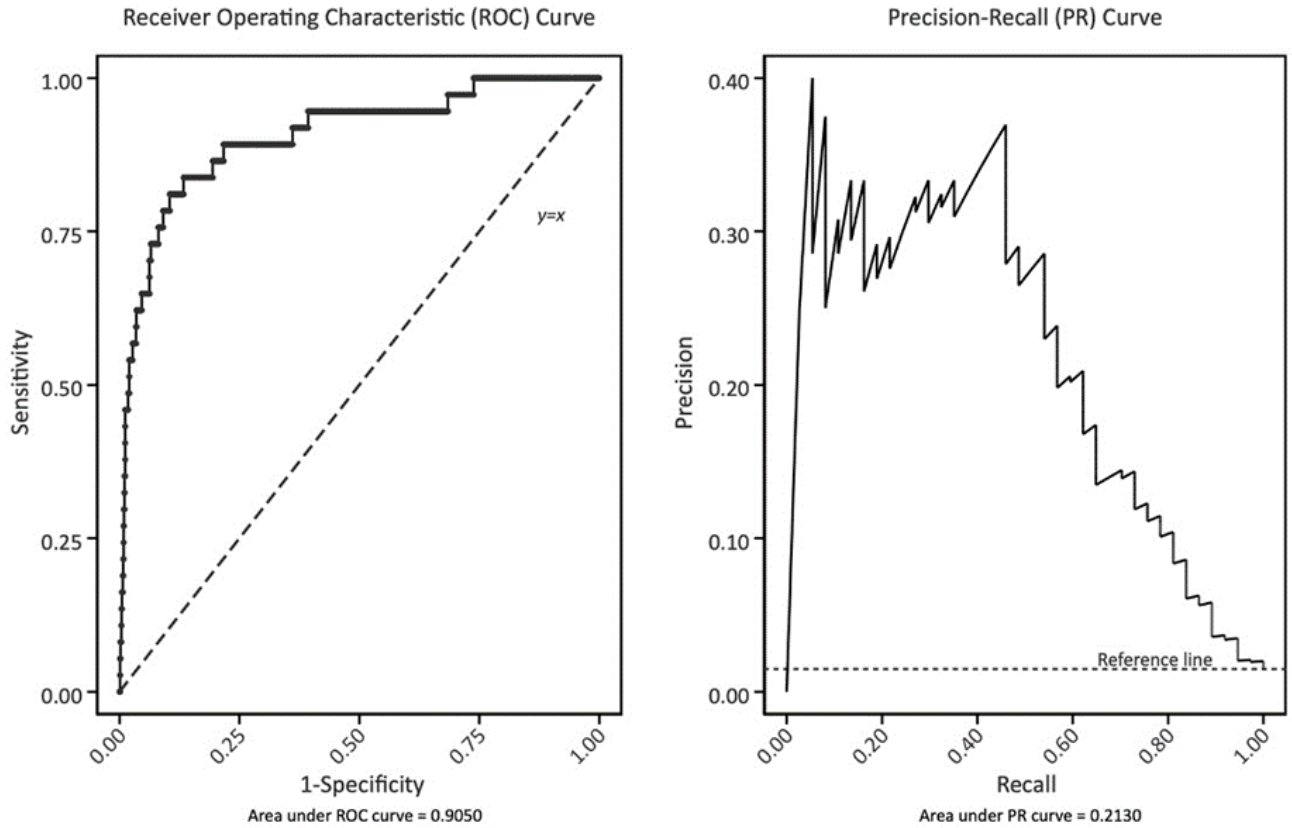
Note: Both curves were generated on the holdout set using coefficient estimates derived from the training set across all payers.

Exhibit 12. ROC Curve and PR Curve; Holdout Set; Medicaid Population



Note: Both curves were generated on the holdout set using coefficient estimates derived from the training set across all payers.

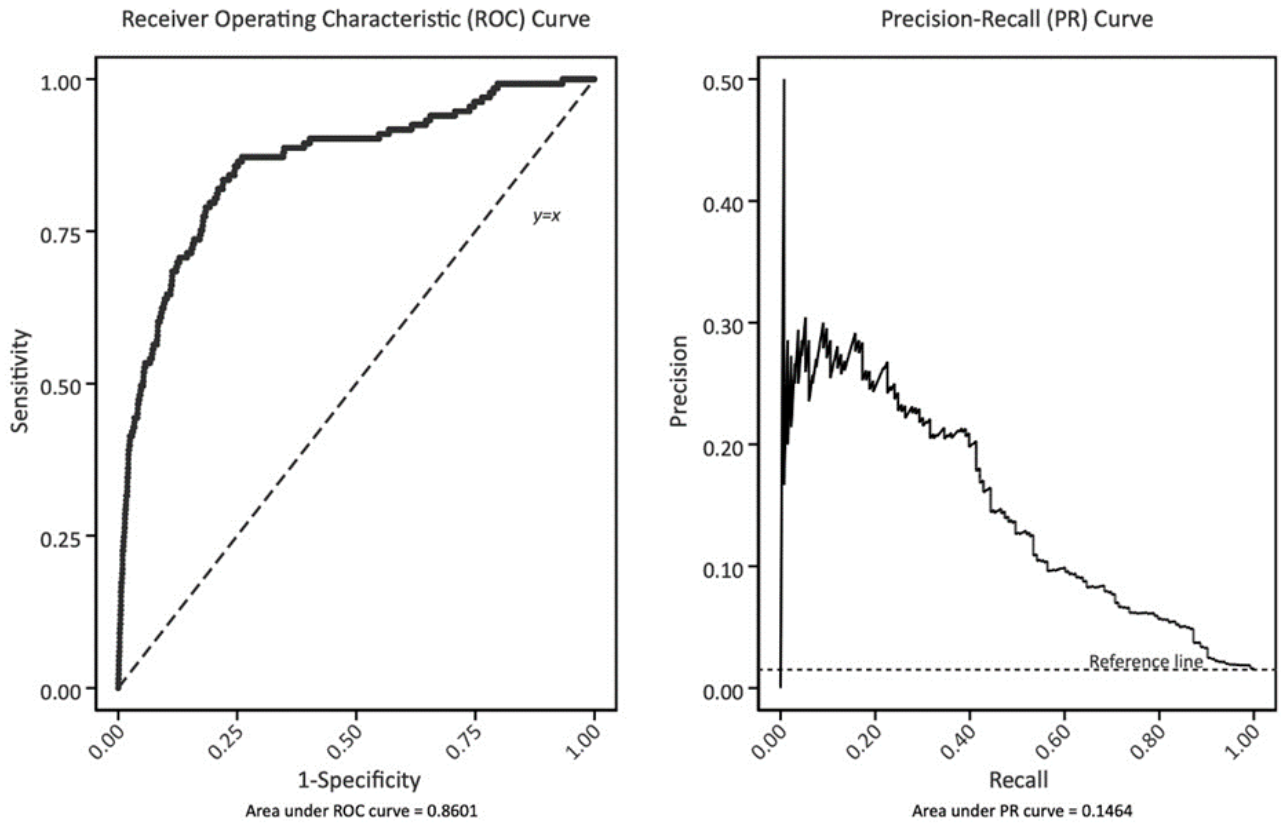
Exhibit 13. ROC Curve and PR Curve; Holdout Set; Commercial Population



Note: Both curves were generated on the holdout set using coefficient estimates derived from the training set across all payers.

We also evaluated the marginal change in model performance by adding social risk factors to the baseline model and **Exhibit 14** shows the ROC and PR curves on the holdout set using coefficient estimates (with the added social risk factors) derived from the training set. No model performance enhancement is detected.

Exhibit 14. ROC Curve and PR Curve; Holdout Set; Added Social Risk Factors



Note: Both curves were generated on the holdout set using coefficient estimates derived from the training set. Social risk factors shown in **Table 37** were included in the logit regression.

Footnote:

[1] Given the prevalence of Medicaid managed care across state, it is of little value to separate out Medicaid FFS from Medicaid managed care.

[Response Ends]

Criterion 3. Feasibility

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in electronic health records (EHRs)

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

Not applicable. This is an eCQM that uses all data elements from defined fields in the electronic health record (EHR).

[Response Ends]

3.05. Complete and attach the NQF Feasibility Score Card.

[Response Begins]

Attached

[Response Ends]

Attachment: 3713e_3713e_AKI NQF feasibility scorecard_vEXTERNAL_To NQF-508.xlsx

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]

During feasibility, a total of 34 sites were evaluated. It was determined that all of the measure's data elements were available within the EHR in structured formats, encoded using nationally accepted terminologies, and aligned with the measure's intent.

However, although the Diagnosis data element (i.e., ICD-10-CM diagnoses used in the measure exclusions[1]) was available within the EHRs, we found the corresponding 'present on admission' (POA) indication, at some sites, was not available for extraction during testing for two primary reasons. The first being a technical issue, where the interface allowing the POA indicator to flow into the EHR was not working and required troubleshooting beyond the testing timeframe. The second was due to the lack of access to the POA indicator data source location during testing. Since POA status is a required element in hospital billing, we do not have concerns about the availability, accuracy and use of standards. Both of these technical issues have been resolved.

Further, we identified that dialysis services was not offered at all prospective test sites. Twenty-nine sites offered dialysis services onsite, of which 7 sites outsourced dialysis services, where the clinical documentation was not available as discrete, structured data. Since the measure is able to capture the intended dialysis population through the use of ICD-10-CM diagnoses codes, there is no concern with these workflow variances.

References:

[1] Hemolytic Uremic Syndrome (HUS), Body Surface Area (BSA) Burns, Traumatic Avulsion, Rapidly Progressive Nephritic Syndrome, Thrombotic thrombocytopenic purpura, Obstetrics or Obstetrics VTE

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),

Attach the fee schedule here, if applicable.

[Response Begins]

There are no fees associated with the use of this eCQM. Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets.

Individuals interested in accessing value set content can request a UMLS license at (https://www.nlm.nih.gov/databases/umls.html).

[Response Ends]

Criterion 4: Use and Usability

4a. Use

4a.01. Check all current uses. For each current use checked, please provide:

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

Not in use

[Not in use Please Explain]

This eCQM is under initial endorsement review and is not currently used in any accountability program. This measure was submitted to the 2022 Measures Under Consideration (MUC) List and will be reviewed by the Measure Applications Partnership (MAP) during their 2022-2023 review cycle.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

[Response Begins]

Public reporting

Payment Program

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

This eCQM is under initial endorsement review and is not currently used in any accountability program. This measure was submitted to the 2022 Measures Under Consideration (MUC) List and will be reviewed by the Measure Applications Partnership (MAP) during their 2022-2023 review cycle. CMS has sought MAP support for implementation in an accountability program (Hospital Inpatient Quality Reporting and Promoting Interoperability Programs) pending feedback received from the MAP, during NQF endorsement, and rulemaking.

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

CMS is seeking MAP's recommendations and support for implementation in the Inpatient Quality Reporting and Promoting Interoperability for eligible hospitals programs.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]

N/A; this measure is being submitted as de novo and has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

For eCQMs included in CMS reporting programs, implementation resources are provided through the CMS eCQI Resource Center and The ONC Project Tracking System (a collaboration platform hosted by the HHS's Office of National Coordinator for Health Information Technology (ONC) that provides users with a common place to transparently log, track, and discuss and clarify issues with eCQM implementation and logic interpretation). As part of the measure rollout, CMS (in

collaboration with The Joint Commission) also provides an annual webinar series for measured entities to review the measure specification, logic, and answer implementation questions.

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

N/A; this measure is being submitted as de novo and has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

N/A; this measure is being submitted as de novo and has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

CMS obtains feedback on all of its measures through various avenues including: (1) Measures Management System (MMS) posting with announcements to stakeholders, (2) NQF endorsement review, (3) Measures Application Partnership (MAP) review, (4) Proposed Rules published in the Federal Register, (5) ongoing feedback from the user community through the QualityNet portal, (6) ongoing review by a Technical Advisory Panel representing key stakeholders and clinical experts, which will continue to support the measure.

Additionally, for eCQMs included in CMS reporting programs, implementation resources are provided through the CMS eCQI Resource Center and The ONC Project Tracking System (a collaboration platform hosted by the HHS's Office of National Coordinator for Health Information Technology (ONC) that provides users with a common place to transparently log, track, and discuss and clarify issues with eCQM implementation). These implementation feedback are evaluated and, as appropriate, presented during the CMS Annual Update Change Review Process for measure refinements.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

N/A; this measure is being submitted as de novo and has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

However, for eCQMs included in CMS reporting programs, there are measure feedback loops provided through the CMS eCQI Resource Center and The ONC Project Tracking System (a collaboration platform hosted by the HHS's Office of National Coordinator for Health Information Technology (ONC) that provides users with a common place to transparently log, track, and discuss and clarify issues with eCQM implementation). Additionally, eCQMs go through an Annual Update Cycle, which includes the Change Review Process (a mechanism for public comment and suggested measure refinements).

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

While this measure does not have usability information from measured entities, as it is being developed de novo and has not been implemented yet, our team sought input from multiple stakeholder groups throughout the measure development process. We believe in a transparent measure development process, and highly value the feedback received on the measure. During development, a technical expert panel composed of a variety of stakeholders was engaged at various stages of development to obtain balanced, expert input. We also sought measure specification feedback from subject matter experts at the Renal Physicians Association (RPA) and American Society of Nephrology (ASN). Finally, we collected feedback from pilot test sites following measure implementation testing, and the post-implementation testing survey to inquire about the measure's usability and its prospect of field implementation.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

Input received from TEP members was instrumental to the development and specification of this measure. Feedback received from stakeholders and SMEs was also explored during the measure testing process. We have made significant modifications to incorporate a more accurate diagnoses of AKI using clinical data, exclusions of pre-hospitalization acquired AKI, and a robust risk adjustment methodology.

[Response Ends]

4b. Usability

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]

This is a new eCQM and there is no time trend information available regarding facility performance improvement. This eCQM is not currently used in any quality improvement program, but a primary goal of the eCQM is to provide hospitals with performance information necessary to implement focused quality improvement efforts.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

We did not identify any unintended consequences during eCQM development or testing. However, CMS is committed to monitoring this eCQM's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, and other negative unintended consequences for patients.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

No unexpected benefits were noted during eCQM development testing.

[Response Ends]

Criterion 5: Related and Competing Measures

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

Related: Patient Safety Indicator (PSI) 10: Postoperative Acute Kidney Injury Requiring Dialysis Rate (Steward: The Agency for Healthcare Research and Quality (AHRQ); CMIT Ref No. 05021) is a related measure that is used to measure episodes of acute kidney injury requiring dialysis per 1,000 elective surgical discharges for patients ages 18 years and older. Although CMS PSI 90, Patient Safety and Adverse Events Composite (NQF #0531), includes PSI 10, it has a much broader scope than this eCQM, as it is a claims-based measure that also includes 9 other types of hospital-acquired complications.

Competing: None

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

No

[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

There are no competing or related measures that are NQF-endorsed, so question 5.04 should be considered not applicable. Since "N/A" is not an available selection in MIMS, we selected "no" because there are differences between our eCQM and the non-NQF endorsed related measure, PSI 10. PSI 10 measures how often hospitalized patients had renal failure requiring dialysis after having an operation. Additionally, PSI 10 utilizes claims data. In comparison, this new AKI measure measures how often stage 2 or greater AKI occurs in the inpatient hospital setting, whether or not the patient receives dialysis, and is developed as an eCQM for both medical and surgical adult inpatients.

Harmonization between our measure and NQF #0531 is not necessary because the measures are not related (i.e., they do not have the same measure focus or the same target population). NQF #0531 is a composite measure of 10 hospital-acquired complications (Patient Safety and Adverse Events Composite), and the only component that overlaps with the proposed measure is PSI 10 (Postoperative Acute Kidney Injury Requiring Dialysis Rate), for which the outcome is defined as "requiring dialysis." PSI 10 is not an endorsed measure, and it has a much narrower focus than the proposed measure, which captures both stage 2 and 3 AKI. Renal replacement therapy is only required for a small subset (circa 3-5%) of patients with a first hospitalization with AKI (https://adr.usrds.org/2020/chronic-kidney-disease/5-acute-kidney-injury), so these are very different outcomes.

Although both NQF #0531 and the proposed AKI measure have a target population of hospitalized adults, their specific denominators are quite different. NQF #0531 is the CMS claims-based version of PSI 90, so its denominator is limited to adult (fee-for service) Medicare beneficiaries, whereas the proposed measure applies to adults of all ages and payers. In addition, the proposed measure is applicable to BOTH medical and surgical patients without EITHER chronic kidney disease (CKD stage 3a or greater) or AKI at presentation, whereas the PSI 10 component of NQF #0531 is focused entirely on surgical patients.

Reference:

AHRQ. (2022). Patient Safety Indicator 10 (PSI 10) Postoperative Acute Kidney Injury Requiring Dialysis Rate ICD-10-CM/PCS Specification v2022

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

We identified no competing measures. This new measure serves as an additional measure to identify how often AKI occurs in the inpatient hospital setting. Although there are many occurrences of AKI in hospital settings, many of which are preventable, there is currently no measure in a Centers for Medicare & Medicaid Services (CMS) quality reporting program or public reporting that quantifies how often AKI occurs in hospitalized patients.

[Response Ends]